UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE AGRONOMÍA

DISEÑO Y ANÁLISIS DE EXPERIMENTOS

FUNDAMENTOS Y APLICACIONES EN AGRONOMÍA

2ª. Edición
Revisada y Ampliada

Ezequiel Abraham López Bautista
Byron Humberto González Ramírez

Docentes Investigadores

GUATEMALA, JUNIO DE 2016

Autores



Ezequiel López es actualmente Profesor Titular VI en las Subáreas de Métodos de Cuantificación e Investigación y de Ejercicio Profesional Supervisado, de la Facultad de Agronomía, USAC. Agrónomo egresado de la Escuela Nacional Central de Agricultura –ENCA- (1991), Ingeniero Agronómo en Sistemas de Producción Agrícola graduado en la Facultad de Agronomía de la Universidad de San Carlos de Guatemala (1999), realizó estudios de Maestría en Agronomía (2002) y de Doctorado en Ciencias (2014), ambos en el Área de concentración Estadística y Experimentación Agronómica en la Escuela Superior de Agricultura "Luiz de Queiroz" de la Universidad de São Paulo. Se ha desempeñado como docente de Estadística Aplicada en Agronomía y ciencias afines, en pre y postgrado en la FAUSAC y en diferentes Universidades Privadas de Guatemala. Además como asesor de trabajos de investigación y consultor estadístico en proyectos de investigación. Su área de interés incluye: diseño y análisis de experimentos aplicados en Agronomía y ciencias afines, métodos estadísticos multivariados, análisis de regresión, geoestadística y aplicación de modelos lineales mixtos en experimentación agronómica.



Byron González es actualmente Profesor Titular VI y Director del Centro de Telemática –CETE- de la Facultad de Agronomía, USAC. Agrónomo egresado de la Escuela Nacional Central de Agricultura –ENCA- (1991), Ingeniero Agronómo en Sistemas de Producción Agrícola graduado en la Facultad de Agronomía de la Universidad de San Carlos de Guatemala (1999), realizó estudios de Maestría en Administración de Empresas en la Universidad Rafael Landívar de Guatemala, actualmente se encuentra en la fase final del Doctorado en Investigación Social de la Universidad Panamericana de Guatemala. Se ha desempeñado como docente de Informática, Estadística Aplicada en Administración, Economía y Agronomía, en pregrado en la FAUSAC y posgrado en diversas Universidades Privadas de Guatemala. Además como asesor de trabajos de investigación y consultor estadístico en proyectos de investigación. Su área de interés incluye: diseño y análisis de experimentos aplicados en Agronomía y ciencias afines, métodos estadísticos aplicados en el Control de la Calidad, y modelación de regresión.

Consultas y sugerencias:

Oficina C–20, Edificio T-8, Universidad de San Carlos de Guatemala, Facultad de Agronomía, Ciudad Universitaria zona 12. Ciudad de Guatemala.

e-mails: lopez_ezequiel@usac.edu.gt
byrong@usac.edu.gt

Dedicatoria

El tiempo y esfuerzo dedicado en escribir, editar y elaborar este documento, ha sido cuantioso. El fuerte deseo de contribuir a mejorar constantemente la enseñanza de la Estadística Experimental a nivel universitario en Guatemala, nos ha brindado las fuerzas para avanzar y poder concluir este material educativo.

Esta obra está dedicada a la memoria de nuestro profesor, compañero y amigo de muchas jornadas, Ing. Agr. M.C. Víctor Manuel Álvarez Cajas.

Agradecimientos

Deseamos expresar nuestro agradecimiento más sincero a:

Los profesores de Estadística de la Facultad de Agronomía (FAUSAC): M.Sc. Marino Barrientos y M.C. Víctor Álvarez Cajas (QEPD); quienes nos acompañaron desde nuestros inicios como Auxiliares de Cátedra en la Subárea de Métodos de Cuantificación e Investigación.

Los profesores del Programa de Estadística y Experimentación Agronómica de la ESALQ/USP: Dra. Clarice Garcia Borges Demétrio, Dr. Décio Barbin, Dr. Silvio Sandoval Zocchi, Dr. Carlos Tadeu Dos Santos Dias y Dra. Sônia Maria Di Estéfano Piedade, por los conocimientos transmitidos.

Los Doctores Mario Melgar y José Luis Quemé del Centro de Investigación y Capacitación de la Caña de Azúcar (CENGICAÑA), por la amistad y motivación para seguir adelante.

Los estudiantes de los cursos de Estadística Aplicada (Agrícola y Forestal) de la FAUSAC y de Diseño y Análisis de Experimentos de la carrera de Ingeniería Forestal del Campus "San Pedro Cláver" de la Universidad Rafael Landívar, en Chamelco, Alta Verapaz; que con sus dudas en clase o fuera de ella, al utilizar las versiones iniciales de este documento, han contribuido a enriquecerlo.

PRESENTACIÓN

El propósito de la Estadística es ofrecer una base objetiva para el análisis de datos observados o medidos, los cuales están sujetos a la variación del azar. Estos datos, en muchos casos, son provenientes de estudios realizados a propósito y en condiciones previamente especificadas y controladas (experimentos). El estudio de los experimentos y sus etapas: planeación, ejecución, toma y análisis de datos es lo que constituye el objetivo de la Estadística Experimental.

La Estadística Experimental en el mundo moderno es una necesidad real, está presente en todas las áreas del conocimiento humano como una herramienta para auxiliar las decisiones a ser tomadas, permitiendo mayor seguridad a la investigación a ser transformada en tecnología y utilizada por la humanidad o por parte de ella.

El interés original de escribir este documento es describir de manera introductoria los conceptos fundamentales del análisis estadístico de experimentos e ilustralos con ejemplos numéricos aplicados en las ciencias agrícolas. Y así poner a disposición del alumno notas que reemplacen a las suyas, para evitar que estén anotando, porque muchas veces no prestan atención por estar preocupados en escribir. El material no pretende reemplazar a los libros, aunque algunas veces las circunstancias han llevado a situaciones de este tipo. Otro de los objetivos es que los alumnos, docentes e investigadores dispongan de un material actualizado que les permita estudiar, repasar y practicar, con ejemplos reales, producto de investigaciones realizadas en su mayor parte en Guatemala, tanto en tesis de grado de la Facultad de Agronomía de la Universidad de San Carlos, como por otras instituciones nacionales e internacionales.

En este documento se recopila información bibliográfica muy valiosa de estadísticos de renombre en América Latina, como el Dr. Ángel Martínez Garza y Dr. Mario Miguel Ojeda (México), Dr. Frederico Pimentel Gomes, Dra. Maria Cristina Stolf Nogueira, Dr. Decio Barbin, Dr. Carlos Tadeu dos Santos Dias y Dra. Clarice García Borges Demétrio (Brasil), Dr. Raúl Macchiavelli (Puerto Rico), Grupo Infostat de la Universidad Nacional de Córdoba (Argentina), entre otros; también de autores clásicos de diseños experimentales (Montgomery, Kempthorne, Steel y Torrie, Ostle, Cochran y Cox, por ejemplo). Así como algunas anotaciones propias, producto de nuestra experiencia adquirida como docentes e investigadores y en los estudios de postgrado. Este texto incluye también programas y salidas proporcionadas por el *software* SAS v. 9.13 (SAS Institute, Cary, NC).

Esta segunda edición contiene más ejercicios prácticos e interpretación de los mismos, una ampliación de los capítulos 4 y 8 referentes a verificación de los supuestos experimentos fundamentales del Análisis de Varianza y Experimentos Factoriales, respectivamente, y más referencias bibliográficas actualizadas. Así mismo hemos incorporado secciones dentro de los capítulos, con instrucciones y salidas del *software* Infostat v. 2016, del cual somos usuarios oficiales en la FAUSAC.

Guatemala, junio de 2016

CONTENIDO

			Página
1.	ASPECTOS GENERALES DEL DISEÑO Y ANÁLISIS DE EXPERIMENTOS		1
	1.1	INTRODUCCIÓN	1
	1.2	PRINCIPIOS BÁSICOS DEL DISEÑO EXPERIMENTAL	10
	1.3	ERROR EXPERIMENTAL	13
	1.4	CARACTERISTICAS DEL DISEÑO EXPERIMENTAL	15
	1.5	ALGUNAS RECOMENDACIONES EN LA APLICACIÓN DE LOS MÉTODOS ESTADÍSTICOS EN LA EXPERIMENTACIÓN	15
	1.6	VALIDEZ EXPERIMENTAL	16
	1.7	RESEÑA HISTORICA Y TENDENCIAS ACTUALES EN EL ANÁLISIS DE DATOS EXPERIMENTALES	18
2.	DISEÑO COMPLETAMENTE AL AZAR		19
	2.1	INTRODUCCIÓN	19
	2.2	ANÁLISIS ESTADÍSTICO	20
	2.3	ANÁLISIS DE VARIANZA PARA UN EXPERIMENTO REALIZADO EN UN DISEÑO COMPLETAMENTE AL AZAR DESBALANCEADO	28
	2.4	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR	30
	2.5	INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR	31
	2.6	ANÁLISIS DE VARIANZA PARA UN EXPERIMENTO REALIZADO MEDIANTE UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO	33
	2.7	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO	37
	2.8	INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO	38
	2.9	EJERCICIOS PROPUESTOS	40
3.	PRUEBAS DE COMPARACIÓN MÚLTIPLE DE MEDIAS		47
	3.1	INTRODUCCIÓN	47
	3.2	COMPARACI0N MÚLTIPLE DE MEDIAS SEGÚN EL CRITERIO DE TUKEY	47
	3.3	COMPARACIÓN MÚLTIPLE DE MEDIAS SEGÚN EL CRITERIO DE DUNCAN	52
	3.4	MÉTODO DE DUNNETT	54
	3.5	EJERCICIOS PROPUESTOS	55
	3.6	PROGRAMA EN SAS PARA EL ANÁLISIS POSTANOVA DE UN EXPERIMENTO	56
	3.7	INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO Y APLICACIÓN DE PRUEBAS DE COMPARACIÓN MÚLTIPLE DE MEDIAS	57
4.	SUPUESTOS FUNDAMENTALES DEL ANÁLISIS DE VARIANZA, VERIFICACIÓN Y TRANSFORMACIÓN DE DATOS		58
	4.1	INTRODUCCIÓN	58
	4.2	TRANSFORMACIÓN DE DATOS	60
	4.3	PROGRAMA EN SAS PARA EL ANÁLISIS DE LOS SUPUESTOS DE NORMALIDAD Y DE HOMOGENEIDAD DE VARIANZAS	70
	4.4	EJERCICIOS PROPUESTOS	72
5.	DISEÑO EN BLOQUES COMPLETOS AL AZAR		74
	5.1	INTRODUCCIÓN	74

	5.2	ANÁLISIS ESTADÍSTICO	76
	5.3	EJERCICIOS PROPUESTOS, DBA	79
	5.4	ANÁLISIS POST ANOVA: CONTRASTES	85
	5.5	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO BLOQUES COMPLETOS AL AZAR, CON CONTRASTES ORTOGONALES	93
	5.6	DISEÑO BLOQUES AL AZAR CON DATOS FALTANTES	95
	5.7	ANÁLISIS DE VARIANZA PARA UN DISEÑO BLOQUES AL AZAR CON MUESTREO	100
	5.8	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO BLOQUES COMPLETOS AL AZAR, CON MUESTREO	103
	5.9	EJERCICIOS PROPUESTOS	105
	5.10	EFICIENCIA DEL DISEÑO EN BLOQUES COMPLETOS AL AZAR	110
6.	USO DE LA REGRESIÓN EN EL ANÁLISIS DE VARIANZA: POLINOMIOS ORTOGONALES		112
	6.1	INTRODUCCIÓN	112
	6.2	MODELOS DE REGRESIÓN POLINOMIAL	112
	6.3	EJEMPLO DE APLICACIÓN	114
	6.4	PROGRAMA EN SAS PARA EL ANÁLISIS DE REGRESIÓN POLINOMIAL EN UN EXPERIMENTO	119
	6.5	EJERCICIOS PROPUESTOS	120
7.	DISEÑO CUADRADO LATINO		123
	7.1	INTRODUCCIÓN	123
	7.2	ANÁLISIS ESTADÍSTICO	125
	7.3	DISEÑO CUADRADO LATINO (DCL) CON DATOS FALTANTES	129
	7.4	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO CUADRADO LATINO	131
	7.5	DISEÑO CUADRADO LATINO (DCL) CON MUESTREO	132
	7.6	PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO CUADRADO LATINO CON MUESTREO	136
	7.7	EJERCICIOS PROPUESTOS	138
8.	EXPERIMENTOS FACTORIALES		142
	8.1	INTRODUCCIÓN	142
	8.2	EXPERIMENTOS FACTORIALES EN ARREGLO COMBINATORIO	147
	8.3	EXPERIMENTOS FACTORIALES CON ARREGLO EN PARCELAS DIVIDIDAS	171
	8.4	EXPERIMENTOS FACTORIALES CON ARREGLO EN PARCELAS SUBDIVIDIDAS	190
	8.5	EXPERIMENTOS TRIFACTORIALES EN ARREGLO COMBINATORIO	199
	8.6	EXPERIMENTOS FACTORIALES CON TRATAMIENTOS ADICIONALES (TESTIGOS) EN INFOSTAT V. 2016 Y MS EXCEL 2013	210
9.	EXPERIMENTOS EN FRANJAS		222
	9.1	INTRODUCCIÓN	222
	9.2	MODELO ESTADÍSTICO-MATEMÁTICO	223
	9.3	HIPÓTESIS	223
	9.4	ESQUEMA DEL ANÁLISIS DE VARIANZA	223
	9.5	EJEMPLO DE APLICACIÓN	225
	9.6	PROGRAMA EN SAS PARA UN EXPERIMENTO BIFACTORIAL CON ARREGLO EN FRANJAS DIVIDIDAS	228
	9.7	EJERCICIOS PROPUESTOS	230
10.	ANÁLISIS DE COVARIANZA (ANCOVA)		232
	10.1	INTRODUCCIÓN	232
	10.2	SUPOSICIONES BÁSICAS DEL ANÁLISIS DE COVARIANZA	232
	10.3	EJEMPLO DE APLICACIÓN	234
	10.4	PROGRAMA EN SAS PARA UN EXPERIMENTO DONDE SE UTILIZÓ ANÁLISIS DE COVARIANZA SIMPLE	239

10.5 EJERCICIOS PROPUESTOS 240
11. ANÁLISIS DE GRUPOS DE EXPERIMENTOS 245
11.1 INTRODUCCIÓN 245
11.2 EJEMPLO DE APLICACIÓN 246
11.3 PROGRAMA EN SAS PARA UNA SERIE DE EXPERIMENTOS 251
11.4 EJERCICIOS PROPUESTO 256
BIBLIOGRAFIA 259
APENDICE 264

CAPÍTULO 1
ASPECTOS GENERALES DEL DISEÑO Y ANÁLISIS DE EXPERIMENTOS

1.1 INTRODUCCIÓN

Los investigadores realizan experimentos virtualmente en todos los campos del saber, por lo general para descubrir algo acerca de un proceso o sistema en particular (Montgomery, 2004). Literalmente, un experimento es una prueba o ensayo. Un experimento diseñado es una prueba o serie de pruebas en las cuales se inducen cambios deliberados en las variables de entrada de un proceso o sistema, de manera que sea posible observar e identificar las causas de los cambios en la respuesta de salida. Un experimento, también es definido como un procedimiento que le permite al investigador, reproducir bajo condiciones "controladas" un fenómeno real con el objeto de obtener la información necesaria para la contrastación objetiva de hipótesis relativas al efecto de factores específicos de la producción.

Existen diversas razones que hacen imperiosa la necesidad de realizar investigación agrícola. Desde un punto de vista muy simple, se podría partir de la necesidad de satisfacer una demanda de alimentos, cada día mayor, tanto en cantidad como en calidad. Los fenómenos biológicos, al no ser determinísticos, requieren de la inferencia estadística para determinar el grado de incertidumbre que existe en los resultados y poder llegar a conclusiones y recomendaciones válidas. En la Agronomía a diferencia de otras ciencias, debido a los efectos de la variabilidad genética y ambiental, es necesario realizar investigación de carácter local.

El conjunto de aspectos que deben resolverse antes de la realización de un experimento son una clara evidencia de que la etapa de planificación de cualquier investigación experimental no debe ser olvidada. Antes de realizar un experimento hay que planificarlo de tal manera que permita obtener la información pertinente al problema bajo investigación, a esta etapa se le conoce como **diseño del experimento**, y puede concebirse como la secuencia completa de pasos a realizar para asegurar que se obtendrá la información necesaria para el contraste de la (s) hipótesis planteada (s). Nogueira (2007) indica que una investigación planificada consiste en las siguientes etapas que dependen de un perfecto entendimiento entre el investigador y el estadístico.

a) Enunciado del problema con formulación de hipótesis (no en términos estadísticos)

Estas hipótesis son primero formuladas en términos científicos dentro del área de estudio (hipótesis científica), y enseguida deben ser expresadas en términos estadísticos (hipótesis estadística). Debe haber una correspondencia perfecta entre las hipótesis científica y estadística para evitar ambigüedades. Por tanto, en el enunciado del problema, la hipótesis científica debe ser formulada de manera precisa y objetiva.

b) Selección de los factores y sus respectivos niveles

En una investigación experimental, generalmente su objetivo es observar de que manera una o más condiciones impuestas puedan interferir en el comportamiento de variables importantes dentro del contexto de la investigación. Estas condiciones impuestas y distintas presentes en un experimento son denominadas: **factores**.

En otras palabras, un factor es cada una de las variables independientes cuyo efecto se está interesado en evaluar. Si un experimento consta de un solo factor se llama experimento **simple**, y si incluye dos o más factores se llama experimento **factorial**.

Los **niveles** de un factor pueden ser definidos como las alternativas de ese factor. Por ejemplo, si se está interesado en estudiar el efecto del factor frecuencias de riego en el rendimiento y evapotranspiración del maíz *(Zea mays)* en la unidad de riego San Cristóbal Acasaguastlán, El Progreso. Las frecuencias: 8, 10, 12 y 14 días, son los niveles (o **tratamientos**) de este factor.

Si el interés fuera diseñar un experimento para estudiar el efecto de 4 niveles de nitrógeno (0, 50, 100 y 150 kg/ha/año) y 4 niveles de fósforo (0, 40, 80 y 120 kg/ha/año) en el rendimiento de cardamomo *(Elettaria cardamomun* M.) en la Serie de Suelos Tamahu, en aldea Choval, Cobán, Alta Verapaz; se tendrían dos factores: nitrógeno con 4 niveles y fósforo con 4 niveles. Se puede decir también que este experimento involucra 16 tratamientos, correspondientes a las combinaciones de niveles de los dos factores considerados.

En cuanto al tipo de nivel, un factor podrá ser cualitativo o cuantitativo. Por otra parte, los factores pueden, de acuerdo a la manera como fueron seleccionados los niveles que los componen, ser considerados como un **factor de efecto fijo o de tipo I** o como un **factor de efecto aleatorio o de tipo II**.

Los factores serán considerados de efecto fijo cuando sus niveles usados por el experimentador son de interés específico. Esto implica que las inferencias estadísticas que se hacen acerca de estos factores se restringen a los niveles específicos estudiados. En general, cuando se trabaja con un efecto fijo, se dice que el espacio inferencial del experimento es el conjunto específico de los niveles de los factores estudiados.

Por otra parte, en algunas situaciones experimentales, los niveles de los factores se eligen al azar de una población mayor de niveles posibles, y el experimentador desea obtener conclusiones acerca de la población completa de los niveles, no sólo de los que se usaron en el experimento. En esta situación se dice que se trata de un factor aleatorio. En el ejemplo b.1 se dice que el factor "variedades" es fijo, y en el ejemplo b.2 el factor "variedades" es aleatorio.

Ejemplo b.1 Un investigador tiene 8 variedades de caña de azúcar para realizar un experimento de evaluación de variedades promisorias.

Ejemplo b.2 Un investigador, entre todas las variedades de caña de azúcar disponibles en el Programa de Mejoramiento Genético, sorteó 6 variedades y con ellas desea realizar un experimento de selección de variedades.

Un experimento factorial podrá existir básicamente tres tipos básicos de relación entre los factores, generando los **factores cruzados**, **factores jerárquicos** y los **mixtos**. De esta forma, los factores pueden ser clasificados en:

Clasificación cruzada: cuando los factores son todos cruzados, y es posible estudiar las interacciones, una vez que al repetirse el mismo nivel del factor en cada uno de los niveles del otro factor es posible estudiar el comportamiento diferencial de un factor cuando el otro cambia de nivel.

Clasificación jerárquica: cuando al cambiar el nivel de un factor (factor de agrupamiento principal) cambian también los niveles del otro factor (factor de subagrupamiento). Se nota, en estos casos, que existe una jerarquía entre los factores, o sea, existe un factor principal y el otro factor está "anidado" o dentro de él. Se observa que el factor de subagrupamiento no puede ser considerado sin que se indique en cual nivel del factor principal está anidado. En este caso no es posible estudiar la interacción entre los factores, pues cuando se cambia de nivel de un factor, los niveles del otro factor pasan a ser otros, imposibilitando evaluar ese tipo de efecto.

Las ideas de interacción o anidamiento, si fuera el caso, son de gran importancia en el proceso de inferencia de datos de experimentos que involucran más de un factor.

Clasificación mixta: surge cuando, en el mismo conjunto de datos, aparecen factores cruzados y jerárquicos. O bien cuando se tienen factores de efectos fijos y aleatorios, además de la media y el error experimental.

Por el propio concepto de factor, se observa que en experimento, la selección de los factores y de sus respectivos niveles (que harán parte del experimento), es básicamente un problema del investigador. En cursos avanzados de Estadística Experimental trataremos el tema de modelos de efectos aleatorios, mixtos y con clasificación jerárquica.

c) Selección de la unidad experimental

En una gran cantidad de situaciones prácticas, la unidad experimental es determinada por la propia naturaleza del material experimental, se llama parcela a la unidad experimental usada y que recibirá un tratamiento. En otras palabras, puede ser considerada como la menor subdivisión del material experimental de tal forma que cualquier unidad experimental podrá recibir tratamientos distintos.

Es interesante observar la diferencia entre unidad experimental y unidad de observación. A las unidades experimentales se les aplican los tratamientos, y una unidad de observación es aquella unidad realmente observada dentro del experimento y que corresponde a una fracción de la unidad experimental, por ejemplo: en experimentos de campo es usual considerar el efecto de borde, en este caso, la unidad de observación es considerada como el área útil (parcela neta).

La unidad experimental puede asumir las más diferentes formas y tamaños, dependiendo del objetivo del trabajo y del material experimental.

Con relación al tamaño de la unidad experimental, Barbin (2013) recomienda que el estadístico deba hacerse las siguientes preguntas:

► ¿Se han realizado experimentos con la especie que va a investigar?

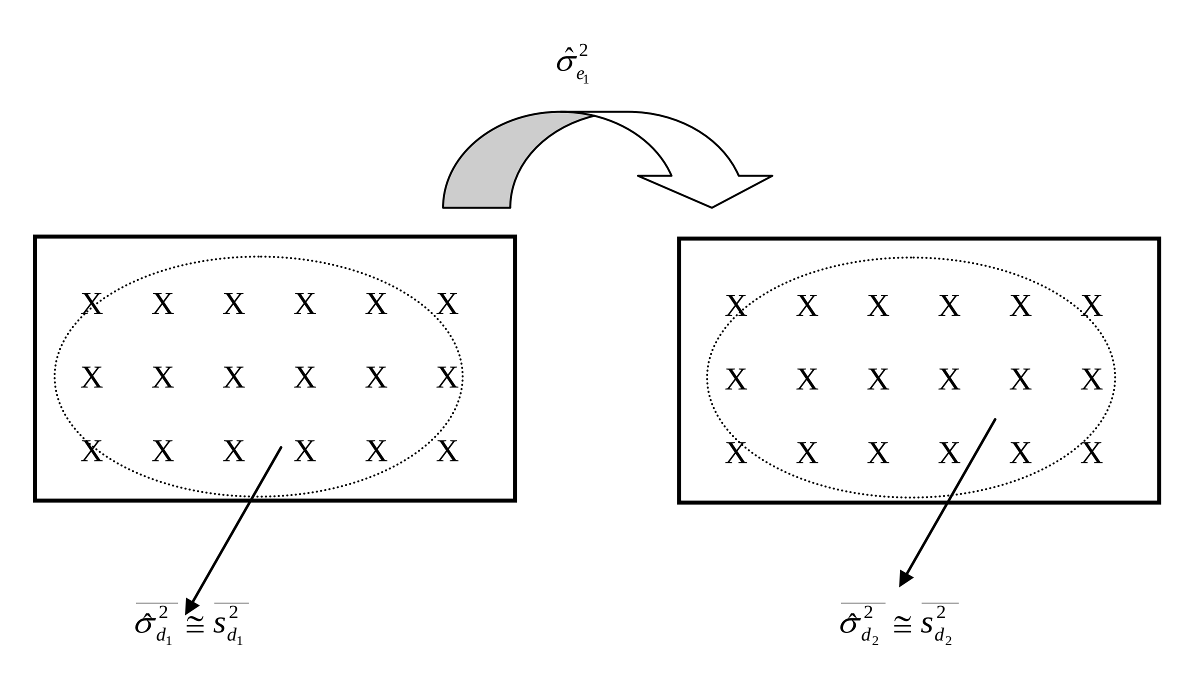
► Si se han realizado, y los resultados fueron satisfactorios, con un coeficiente de variación bajo, se debe trabajar con el mismo tamaño de parcela.

► Si no se han realizado, se debe preguntar al investigador sobre la homogeneidad del material experimental.

  a) Si es homogéneo: las parcelas pueden ser pequeñas.

  b) Si es heterogéneo: las parcelas deben ser grandes, pero, de acuerdo con la disponibilidad del material experimental y del personal capacitado para trabajar con ese material.

Cuando existe duda sobre el tamaño de la unidad experimental, el investigador debe ser orientado en el sentido de ofrecer al estadístico, o aquel que irá hacer los análisis estadísticos, las mediciones individuales, pues, con eso se puede medir la VARIACIÓN DENTRO Y ENTRE LAS UNIDADES EXPERIMENTALES. Esto dará una orientación sobre el tamaño de la unidad experimental. Por ejemplo, tomando el caso de dos unidades experimentales:



$\overline{\hat{\sigma}^2_d} \cong \overline{s^2_d}$ = varianza estimada de las muestras dentro de las unidades experimentales (está dada por el error de muestreo.

$\hat{\sigma}^2_e$ = Varianza estimada entre unidades experimentales (dada por el error experimental)

Algunas conclusiones importantes:

- $\overline{\hat{\sigma}^2_d} > \overline{\hat{\sigma}^2_e}$: Se debe aumentar el tamaño de la unidad experimental.
- $\overline{\hat{\sigma}^2_d} < \overline{\hat{\sigma}^2_e}$: Se puede mantener o aumentar el tamaño de la unidad experimental.

EJEMPLO 1.1

Considere los siguientes resultados de un experimento que se realizó con submuestreo:

Cuadrado medio del error de muestreo (s^2_d): 7.93
Cuadrado medio del error experimental (s^2_e): 11.278
Número de muestras (m): 3

Con esta información, encuentre el valor de σ^2_e y discuta los resultados.

Solución:

$$s^2_e = \sigma^2_d + m \times \sigma^2_e$$

Como σ^2_d puede ser estimada a través de s^2_d, se tiene que: $s^2_e = s^2_d + m \times \sigma^2_e$
Y el valor de σ^2_e se obtiene así:

$$\sigma_e^2 = \frac{s_e^2 - s_d^2}{m} = \frac{11.278 - 7.93}{3} = 1.2$$ (Varianza real entre unidades experimentales)

Conclusión:

Como $\sigma_d^2 > \sigma_e^2$: se plantea la necesidad de aumentar el tamaño de las unidades experimentales, y la precisión del muestreo.

Chacín (1976) indica que los factores que influyen en el tamaño y la forma de la parcela:

1 Extensión superficial del terreno disponible
2 Tipo de suelo
3 Clase de cultivo
4 Objetivo de la investigación
5 Uniformidad del material experimental
6 Método de cultivo
7 Número de tratamientos y de repeticiones
8 Recursos económicos disponibles
9 Muestreo dentro de la unidad experimental
10 Grado de precisión deseado.

Chacín (1976) también cita que el método de la **Máxima Curvatura** fue el primero que se utilizó para conseguir tamaño de parcela, es el método general; hace uso de los llamados ensayos en blanco que consisten en sembrar un área relativamente grande de un solo cultivo aplicando las mismas técnicas, los mismos métodos culturales, tratando de que exista la mayor uniformidad posible; por ello se denominan **ensayos de uniformidad,** ya que el único factor que se desea que varíe es la heterogeneidad del suelo.

E1 área es cosechada en unidades básicas (área pequeña de cosecha en la cual es dividida el área total), a estas unidades se les calcula posteriormente la varianza, la desviación estándar, la media y el coeficiente de variación. Estos también son calculados para parcelas más grandes creadas con la unión de unidades básicas contiguas. Luego se plantean en una curva la desviación estándar con los diferentes tamaños de parcela obtenidas, también se puede utilizar el coeficiente de variación. Se localiza entonces el punto de curvatura máxima y la consiguiente estimación del tamaño óptimo de parcela en forma gráfica. E1 punto donde se encuentra el tamaño óptimo es aquel donde un aumento en "X" (área) no produce un descenso marcado en el coeficiente de variación, localizándose el punto de máxima curvatura. Para mayor información sobre el procedimiento, consulte las tesis de Barrientos García, M. (1981), Álvarez Cajas, VM (1982) y Hernández Dávila, A. (1982).

En cuanto a la forma de la parcela, Chacin (1976) afirma que tiene menos influencia que el tamaño en la disminución de los coeficientes de variación, por lo tanto en la detección de diferencia entre tratamientos. Sin embargo, algunas veces la forma de parcela puede ser muy importante, ella depende mucho del manejo de las diferentes prácticas culturales, forma general del campo y exigencia del cultivo que se trate.

Existe una gran variación en la forma de la parcela, puede haber parcelas rectangulares de diferentes dimensiones y en diferentes sentidos, al igual que forma de parcela cuadrada. La forma rectangular tiene las ventajas de que muchas veces facilita las prácticas culturales del cultivo, uso de máquina, riego, fertilización, control de plagas, etc. Según Banzatto y Kronka (2011) las parcelas rectangulares son más eficientes en la superación de la heterogeneidad del suelo cuando su eje mayor está en la dirección de la mayor variación del suelo.

A continuación se presenta una recopilación de algunos tamaños de unidad experimental utilizados en experimentos que se han realizado en Guatemala.

Caña de azúcar.

- 6 surcos con caña, de 10 metros de largo, con una distancia de 1.5 m entre surco; haciendo un área de 90 m^2 por unidad experimental.

- 5 surcos con caña, de 10 metros de largo, con una distancia de 1.5 m entre surco; haciendo un área de 75 m^2 por unidad experimental.

Gramíneas (maíz, arroz, sorgo, etc.)

- 3 a 10 surcos de plantas, de 10 m de largo.

Cítricos.

- 1 hasta 4 plantas por parcela

Piña.

- 16 plantas, distanciadas 0.35 m entre planta y 1.6 m entre surco. 2 surcos/unidad experimental de 3.15 m de largo y 2.1 m de ancho (6.62 m^2)

- 20 plantas, distanciadas 0.5 m entre planta y 1.0 m entre surco. 2 surcos/unidad experimental de 3 m de largo y 2 m de ancho. (6 m^2)

- 40 plantas, distanciadas 0.3 m entre planta y 0.4 m entre surco. 3 surcos/unidad experimental de 4 m de largo y 1.2 m de ancho. (4.8 m^2)

Brócoli.

- 132 plantas, distanciadas 0.45 m entre planta y 0.5 m entre surco. 11 surcos/unidad experimental de 5.4 m de largo y 6 m de ancho. (32.4 m^2)

Chile pimiento.

- 18 plantas (en 6.48 m^2), 23 plantas (en 7.68 m^2), 27 plantas (en 8.64 m^2), 48 plantas (en 11.20 m^2), hasta 180 plantas (en 57.6 m^2)

Experimentos en laboratorio.

- 3 semillas germinadas dentro de un tubo de ensayo.
- 1 explante de aproximadamente 5 mm colocado en frascos de vidrio con 25 ml de medio de propagación.
- 1 meristemo con 2 primordios florales en cada tubo de 25 × 150 mm. En total 10 meristemos por unidad experimental.

Finalmente, la selección de la unidad experimental, de un modo general, debe ser orientada en el sentido de minimizar el error experimental, esto es, las unidades experimentales deben ser lo más homogéneas posibles, para que, cuando sean sometidas a tratamientos diferentes, sus efectos sean fácilmente detectados.

d) Selección de las variables a ser medidas en las unidades experimentales.

En los experimentos, los factores y sus niveles son definidos y aplicados, y sus efectos son observados en las variables de interés, denominadas: variables de respuesta (variables dependientes). Su análisis es el objetivo básico de un experimento, una vez que el estudio de la influencia de los factores sobre ellas es el que deberá responder a las preguntas de la investigación.

Las medidas realizadas en las unidades experimentales luego de ser sometidas a los tratamientos, constituyen los valores de la variable de respuesta, la cual en general, es determinada "*a priori*" por el investigador, o sea que, él sabe cuál es la variable que quiere medir. Lo que constituye un problema, algunas veces, es la manera como la variable es medida, pues de esto depende la precisión de las observaciones y la distribución de probabilidad de la variable, lo cual es esencial para la selección del método de análisis estadístico.

Debe seleccionarse la o las variables más idóneas que realmente permitan juzgar o evaluar el comportamiento o efecto de los tratamientos en estudio. Por ejemplo, Castillo Fratti (1996) realizó un experimento para evaluar el efecto de la incisión anular sobre la calidad y el rendimiento de la fruta de uva de mesa en dos localidades del nororiente de Guatemala. Utilizó un diseño de bloques completos al azar, con 5 tratamientos y 4 repeticiones; la unidad experimental estuvo constituida por 3 plantas de uva completamente desarrolladas (adultas) y podadas (en un área de 6 m^2). Las variables de respuesta medidas las clasificó en:

- De rendimiento: número de racimos por planta, peso del racimo individual y promedio por planta, número de bayas por racimo, peso y volumen de la baya.

- De calidad: tamaño de la fruta, grados brix y prueba de aceptación.

- Otras variables: relación de yemas vegetativas y yemas florales y días a la cosecha.

e) Selección del diseño experimental.

Para elegir el diseño es necesario considerar el número de repeticiones, las condiciones del sitio experimental y las condiciones de manejo del ensayo. De acuerdo con Martínez (1994), cuando se proyecta un experimento, el investigador debe tener en mente dos aspectos básicos a saber: (1) la elección, propiamente hablando, del arreglo geométrico o diseño experimental, que tiene por objetivo definir el arreglo de los tratamientos sobre las unidades experimentales, y (2) la composición o proyecto de los tratamientos, lo cual constituye el diseño de los tratamientos, que se refiere al proyecto de las combinaciones de niveles, cuando se examina el efecto de dos ó más factores, sobre una característica en estudio.

f) Análisis estadístico de los datos experimentales

Deben emplearse métodos estadísticos para analizar los datos, de modo que los resultados y conclusiones sean objetivos más que apreciativos. Si el experimento se diseñó correctamente y si se ha realizado conforme al diseño, los métodos estadísticos que se requieren no son complicados. Existen diversos paquetes estadísticos para el análisis de datos, por ejemplo: SAS (Statistical Analysis System, "Sistema de Análisis Estadístico"), R, STATISTICA, SYSTAT, SPSS, INFOSTAT, MSTAT, GENSTAT, GENES, etc.; y varios métodos gráficos sencillos que son importantes en la interpretación de tales datos. En las computadoras del Centro de Telemática (CETE) de la FAUSAC actualmente

disponemos de la licencia del Infostat v. 2016. Este programa también está disponible en versión estudiantil y libre, para obtenerla ingrese a www.infostat.com.ar

Las técnicas estadísticas, aunadas a un buen conocimiento técnico o del proceso y al sentido común, suelen llevar a conclusiones razonables.

Uno de los métodos de análisis de datos provenientes de experimentos es el del análisis de varianza (ANOVA), generalmente atribuido a Ronald Fisher. El análisis de varianza **es un proceso aritmético y estadístico que consiste en descomponer la variación total en fuentes o causas de variación**. La variación total se entiende como la variación entre las unidades experimentales. Este proceso permite la descomposición de la variación y no de la varianza. Para utilizarse correctamente este proceso es fundamental conocer, además de la variación inherente a las unidades experimentales, cuáles son las otras fuentes previstas de variación.

Las fuentes de variación presentes en un análisis de varianza deberán ser los efectos principales, los efectos de interacción, los efectos de anidamiento y el residuo, que estimará la variabilidad inherente a las unidades experimentales.

El efecto principal corresponde al efecto de un factor tomado como una media de los demás factores, incluyendo las repeticiones, presentes en el experimento.

El efecto referente a las interacciones mide el diferencial de un factor cuando el otro cambia de nivel. Cuando involucra más de dos factores su significado es más complejo.

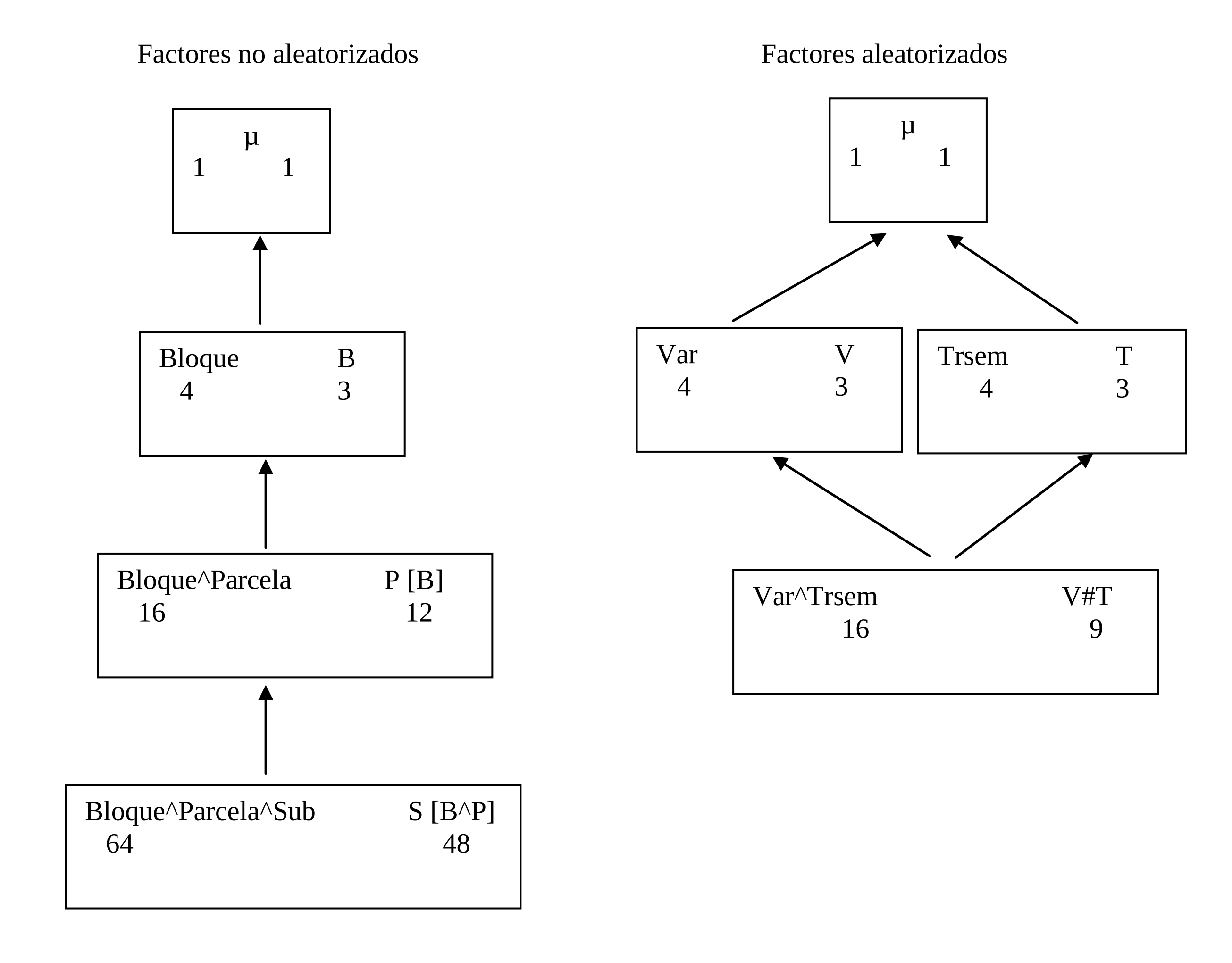
El efecto anidado también es un efecto medio con relación a los demás factores dentro de los cuales él no está anidado.

El residuo es la fuente de variación que representa toda la variación no provocada por el investigador. Si ninguna fuente sistemática estuviera contenida en el residuo, entonces él es un estimador no sesgado de la varianza poblacional de la variable de respuesta.

La identificación de todas las fuentes de variación presentes en un experimento, puede ser fácilmente ser realizado a través de un diagrama esquemático conocido como diagrama de Hasse, que es uma herramienta gráfica que tiene como objetivo facilitar la comprensión de la estructura presente entre los factores experimentales, por muy complejos que sean los ensayos. Además de una mejor visualización del experimento, provee através de reglas propuestas en la literatura, los números de grados de libertad de cada fator. Bajo condición de ortogonalidad del diseño, se pueden obtener tambien las matrices núcleo de las formas cuadráticas para las sumas de cuadrados y las esperanzas de los cuadrados medios, propiciando la relación adecuada para la aplicación de la prueba de F (Alcarde, 2007).

El diagrama de Hasse es un resultado de conceptos matemáticos referidos a estructura lática y diseños de estructura experimental, discutidos por Trockmorton (1961), Kempthorne (1961, 1994), Zyskind (1961) y Kempthorne en asocio con Folks (1961), todos ellos citados por Restrepo, L. (2007).

A continuación se presenta un diagrama de Hasse que contiene los grados libertad, para un experimento bifactorial en bloques al azar con arreglo en parcelas divididas; fueron evaluadas 4 variedades de avena (Var), 4 tratamientos para desinfección de semillas (Trsem), distribuidos en 4 bloques. Los niveles de las variedades fueron colocados en las parcelas grandes y los tratamientos de desinfección de semillas en las parcelas pequeñas. La descripción detallada de este tipo de ensayos se encuentra en el capítulo 8.



Efectos cruzados: se representan así, por ejemplo: V #T, significa todas las posibles combinaciones de los niveles de variedades con los de tratamiento de las semillas.

Efectos aninados: se representan así, por ejemplo: P[B], significa que las parcelas están anidadas dentro de los bloques.

El esquema del ANOVA con los grados de libertad queda de la siguiente manera:

Fuentes de Variación	Grados de Libertad
Bloques	3
Parcelas [Bloques]	12
Variedades	3
Residuo (a)	9
Subparcelas [Bloques ^ Parcelas]	48
Trsem	3
Trsem#Variedades	9
Residuo (b)	36

1.2 PRINCIPIOS BÁSICOS DEL DISEÑO EXPERIMENTAL

1.2.1 Repetición

La repetición se refiere al número de veces que el tratamiento aparece en el experimento, y es determinado a partir de informaciones sobre la variabilidad de las unidades experimentales en términos de la variable dependiente.

La importancia de la repetición está en proporcionar una estimación del error experimental (varianza), aumentar el poder de las pruebas estadísticas como F y las pruebas de médias y aumentar la precisión de las estimaciones de las medias de los tratamientos. El número de repeticiones a ser utilizado es fijado por la experiencia del investigador y dependerá de la grandeza del error deseado y de la variabilidad del material en estudio. En general se tiene como regla práctica, que el número mínimo de parcelas que irá a componer un experimento es de 20 y que el error experimental (residuo) debe tener por lo menos 10 grados de libertad (Gomes, 2000). Una solución para determinar el número de repeticiones es citada por Gomes (2000), y consiste en el uso de la prueba de Tukey y es dada por la siguiente ecuación:

$$\text{Número de repeticiones} = \frac{q^2 \times s^2 \times F_{(n_1,n_2,\alpha)}}{d^2},$$

Siendo:

q = amplitud total estudentizada para el experimento a ser realizado (valor tabulado)

s^2 = es el cuadrado medio del error (o residuo) del experimento anterior, con n_2 grados de libertad.

$F_{(n_1,n_2,\alpha)}$ = valor tabulado para el nivel α de significancia seleccionado.

d = diferencia mínima que deberá ser estadísticamente probada por el experimento.

Es de notar, que el uso de esta ecuación requiere datos de experimentos semejantes. Este número de repeticiones garantizará una probabilidad α de que el ensayo no vaya a comprobar una diferencia d, esto es, una probabilidad $1-\alpha$ de que sea comprobada estadísticamente por la prueba de Tukey. Como los valores de q y de F a ser usados en el segundo miembro dependen del número de repeticiones, es claro que solo se puede obtener una solución por aproximaciones sucesivas, a partir de una tentativa inicial cualquiera. A continuación se presenta un ejemplo ilustrativo.

EJEMPLO 1.2

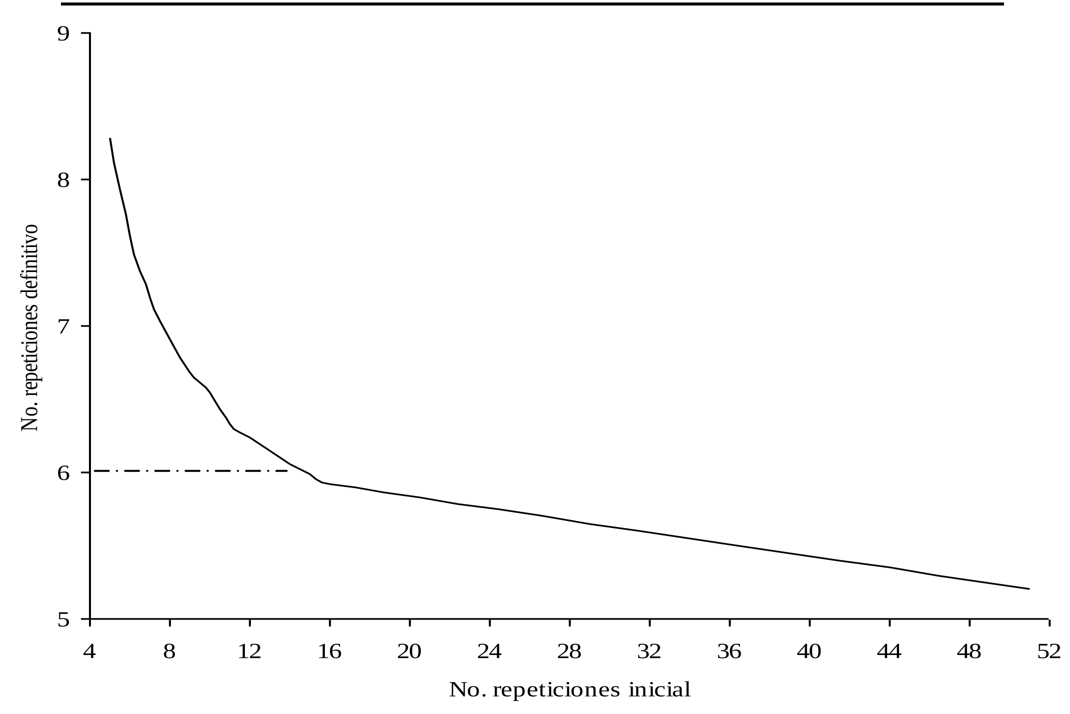
Se planificó un experimento con 5 variedades de caña de azúcar, de un ensayo anterior se obtuvo una estimación de la desviación estándar del residuo (la raíz cuadrada del cuadrado medio del error experimental), $s_2 = 7.44$ TCH (toneladas métricas de caña por hectárea), con $n_2 = 60$ grados de libertad, y considerando que el nuevo experimento deba comprobar por la prueba de Tukey cualquier diferencia de producción de 15 TCH o más. Si en el experimento se utilizará el diseño bloques completos al azar y se considerarán 5 repeticiones, como valor inicial.

En este caso, se tendrán 4 grados de libertad $(t-1)$ para tratamientos (variedades) y $n_1 = 16$ grados de libertad para el residuo (gl residuo = $(t-1)(r-1) = (5-1)(5-1) = 16$), luego, con un nivel de 5% de probabilidad se encuentra en la Tabla 2 que $q = 4.33$. Por otra parte, el valor de F, también con un nivel de significancia de 5% de probabilidad, con $n_1 = 16$ y $n_2 = 60$ grados de libertad, es $F = 1.81$. Con estos datos se tiene que:

$$\text{Número de repeticiones} = \frac{4.33^2 \times 7.44^2 \times 1.81}{15^2} = 8.3$$

Si se continúa con el procedimiento iterativo, se generará la tabla y gráfica siguientes:

r inicial	$\mathbf{n_1}$	$\mathbf{n_2}$	**Q**	**F**	**r definitivo**
5	16	60	4.33	1.82	8.3
6	20	60	4.23	1.75	7.6
7	24	60	4.17	1.70	7.2
8	28	60	4.13	1.66	6.9
9	32	60	4.096	1.64	6.7
10	36	60	4.082	1.61	6.5
11	40	60	4.04	1.59	6.3
12	44	60	4.028	1.58	6.2
13	48	60	4.016	1.57	6.14
14	52	60	4.004	1.55	6.06
15	56	60	3.992	1.54	5.99
16	60	60	3.98	1.53	5.92
51	200	60	3.86	1.44	5.21



A partir de esta gráfica, se puede concluir que el número de repeticiones estaría entre 5 y 6. Gomes, FP (2000) también indica que una solución interesante se obtiene de manera análoga, cuando se conoce el coeficiente de variación (CV) y la diferencia mínima d, en porcentaje, a ser comprobada. Se tiene entonces que:

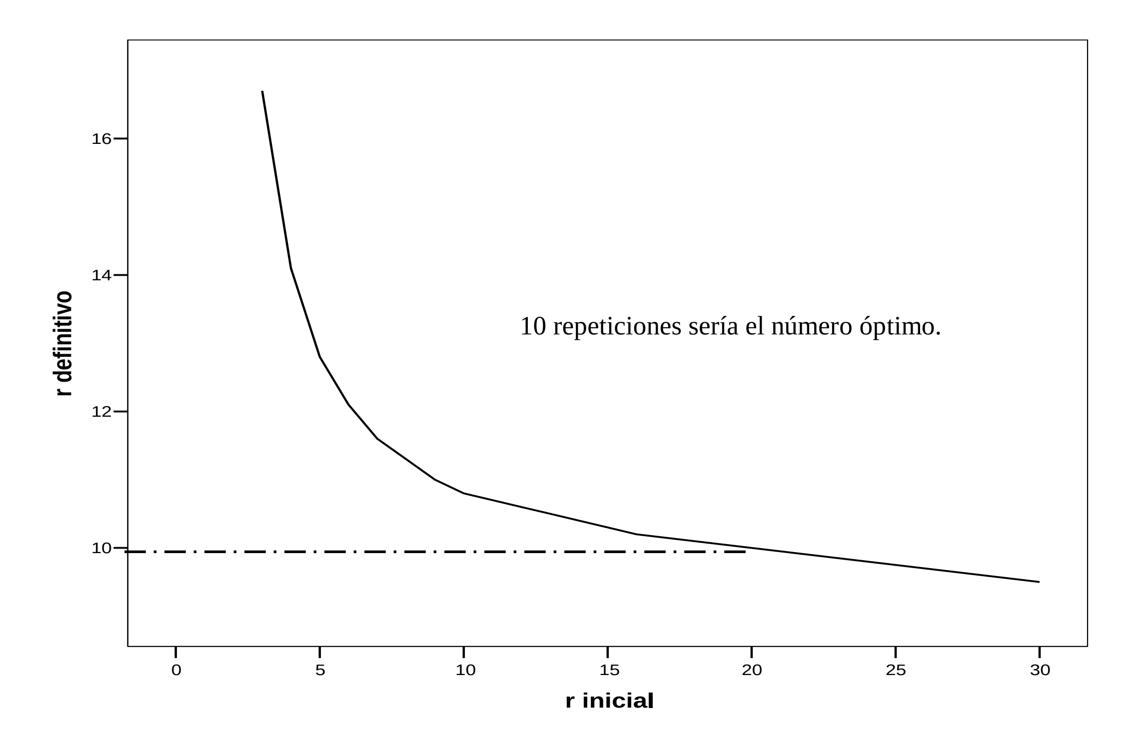
$$\text{Número de repeticiones} = \frac{q^2 \times (CV\%)^2 \times F_{(n_1, n_2, \alpha)}}{(d\%)^2}$$

Si por ejemplo, el caso de un coeficiente de variación, del experimento anterior, CV =15%, calculado con una desviación estándar del residuo que tenía n_2 = 60 grados de libertad. Si se desea saber el número necesario de repeticiones de un experimento a ser realizado, con 8 tratamientos y d =25%, tomando como punto de partida 3 repeticiones, se encuentra:

$$\text{Número de repeticiones} = \frac{4.99^2 \times (15)^2 \times 1.86}{(25)^2} = 16.7$$

Si se continúa con el procedimiento iterativo, se generará la tabla y gráfica siguientes:

r inicial	$\mathbf{n_1}$	$\mathbf{n_2}$	**Q**	**F**	**r definitivo**
3	14	60	4.99	1.86	16.7
4	21	60	4.75	1.73	14.1
5	28	60	4.63	1.66	12.8
6	35	60	4.56	1.62	12.1
7	42	60	4.51	1.59	11.6
8	49	60	4.48	1.56	11.3
9	56	60	4.46	1.54	11.0
10	63	60	4.44	1.53	10.8
11	70	60	4.43	1.52	10.7
12	77	60	4.42	1.51	10.6
13	84	60	4.41	1.50	10.5
14	91	60	4.40	1.49	10.4
15	98	60	4.39	1.48	10.3
16	105	60	4.38	1.48	10.2
30	203	60	4.29	1.44	9.5



1.2.2 Aleatorización

Se refiere a la asignación de los tratamientos a las unidades experimentales, de tal manera que todas las unidades tengan la misma probabilidad de recibir un determinado tratamiento, a través de sorteo o de la tabla de números aleatorios. Aleatorizar los tratamientos significa eliminar tendencias, errores sistemáticos y preferencias que puedan darse en la distribución de los tratamientos a las unidades experimentales. Si la repetición hace posible una prueba de hipótesis, la aleatorización contribuye a hacerla válida. Su función es evitar posibles inducciones en las conclusiones de la investigación y garantizar la independencia de los datos, tornando las estimaciones de las medias de los tratamientos y del error experimental no tendencioso. El proceso de aleatorización de los tratamientos en las unidades experimentales puede ser realizado en los programas SAS y R.

Los principios básicos de repetición y de la aleatorización son obligatorios en cualquier diseño experimental.

1.2.3 Control local

Se refiere a restringir la aleatorización de los tratamientos a grupos de unidades experimentales con poca variabilidad entre sí. Este principio es aplicado cuando se sabe “a priori” de la existencia de variación en el ambiente experimental, no pudiendo correr el riesgo de efectuar una aleatorización completa de los tratamientos en las unidades experimentales. Esta situación puede y es muchas veces resuelta de forma satisfactoria a través de un agrupamiento de unidades experimentales homogéneas entre si y, dentro de esos grupos de unidades homogéneas, también conocido como bloques, es realizada la aleatorización.

Este procedimiento es adoptado con la finalidad de hacer más eficiente el experimento, incrementando la sensibilidad de las pruebas de significancia al reducir la magnitud del error experimental.

1.3 ERROR EXPERIMENTAL

Es un término propio de la experimentación y no es sinónimo de equivocación o descuido sino se refiere a la imposibilidad de poder llegar a resultados idénticos con unidades experimentales tratadas de la misma manera. Existen diferencias debidas a factores genéticos y ambientales que van más allá del control que el investigador pueda ejercer. Al error experimental también se le define como la variabilidad no controlada que existe entre las unidades experimentales que reciben la aplicación del mismo tratamiento. La existencia del error experimental hace necesarias las técnicas estadísticas de análisis de los resultados para obtener buenas estimaciones del efecto de los tratamientos y de las diferencias entre los mismos, las primeras mediante el promedio de los resultados de las unidades experimentales que recibieron cada tratamiento, y el segundo determinando la probabilidad de que las diferencias entre tratamientos hayan ocurrido por casualidad. Se establecen dos fuentes principales de error experimental:

1. Variabilidad inherente al material experimental sobre el cual se aplican los tratamientos. En general, esta variabilidad no es posible evitarla y comprende la heterogeneidad del suelo, la variabilidad genética, la variabilidad del clima, etc.

2. Variabilidad resultante de cualquier falta de uniformidad en la ejecución del experimento, es decir, la deficiencia de poder uniformizar la técnica experimental, tal el caso de: la preparación del suelo, densidad de siembra, prácticas culturales, mediciones y toma de datos, etc.

Según su magnitud, el error experimental producirá un enmascaramiento de los verdaderos efectos de los tratamientos, lo que a su vez producirá un incremento en el riesgo de obtener conclusiones equivocadas. Por esta razón es imperativo reducir (puesto que no es posible eliminar), el efecto de todos aquellos factores que no son lo que interesan en el estudio.

Mayor error experimental equivale a menor precisión[1]. Como un criterio para juzgar la magnitud del error experimental se utiliza el coeficiente de variación, para una misma variable de respuesta, el incremento del coeficiente de variación significa un mayor error experimental, lo cual puede deberse a deficiencias en el diseño o a poco cuidado en la ejecución de las diversas actividades que el experimento conlleva, si se supone que el experimento fue bien diseñado, un coeficiente de variación alto, generalmente significa mal manejo del experimento, (esto depende del tipo de experimento y de la escala utilizada en la medición de la variable respuesta).

Los procedimientos que se pueden emplear para incrementar la precisión de un experimento consisten en tratar de reducir el error experimental, esto puede lograrse por medio de las siguientes acciones:

a) Incrementando el número de repeticiones, aunque el grado de mejoramiento decrece rápidamente cuando el número de repeticiones aumenta, por ejemplo, para duplicar la precisión de un experimento con 4 repeticiones éstas se deberán aumentar a 16.

b) Seleccionando material experimental tan homogéneo como sea posible o estratificación cuidadosa del existente (buscar el material experimental que tenga las características deseadas o seleccionar el diseño más apropiado a las características del material con que se cuente). Es deseable seleccionar material uniforme pero no se debe olvidar de la población acerca de la cual se desea hacer inferencias, razón por la cual en agronomía es recomendable usar material experimental de los tipos que se emplearán en la producción real y estratificar adecuadamente.

c) Ejecutando cuidadosamente el experimento, de manera que se logre uniformidad en la técnica empleada para que se tenga la oportunidad de medir las diferencias entre los efectos de los tratamientos.

d) Seleccionando el tamaño y forma de la unidad experimental. Generalmente a mayor tamaño se conseguirá menor error experimental. Las parcelas rectangulares de terreno deben ubicarse con sus ejes mayores en el sentido de la mayor variabilidad del suelo. En animales es preferible usar un animal como unidad experimental en lugar de tener varios y así obtener mayor número de repeticiones.

e) Usando informaciones proporcionadas por variables aleatorias relacionadas (llamadas covariables), esto se refiere a la toma de datos adicionales, el peso final o la ganancia de peso de un animal depende del peso inicial, el rendimiento depende del número de plantas cosechadas, del contenido de humedad, etc. Esta información adicional es utilizada en el análisis de covarianza.

f) Tomando los datos de variables de respuesta solo en la parcela útil para evitar el efecto de bordes y cabeceras.

g) Utilizando métodos de análisis adecuados al diseño y a la naturaleza de los datos.

[1] Precisión: es la capacidad de detectar diferencias verdaderas entre los promedios de los tratamientos

1.4 CARACTERÍSTICAS DEL DISEÑO EXPERIMENTAL

"Un buen diseño experimental es aquel que proporciona la información requerida con el mínimo esfuerzo experimental". La información requerida se refiere a que los datos permitan un análisis objetivo que conduzca a conclusiones válidas con respecto al problema que se estudia, en cuanto al esfuerzo experimental se entiende por el ahorro de tiempo, dinero, personal y material experimental. Vale hacer énfasis en características como:

1. **Simplicidad**: la selección de los tratamientos y su disposición en el experimento debe ser lo más simple posible, pero consistente con los objetivos del problema.

2. **Grado de precisión**: el experimento debe ser capaz de medir diferencias entre tratamientos, con el grado de precisión deseado, lo cual está muy asociado a diseño empleado, al tamaño de la unidad experimental y al número de repeticiones.

3. **Ausencia de errores sistemáticos**: las unidades experimentales que reciben el mismo tratamiento no deben tener diferencias sistemáticas con las que reciben cualquier otro tratamiento, para poder obtener una buena estimación del efecto de los tratamientos.

4. **Amplio rango de validez de las conclusiones**: es deseable tratar de que las conclusiones a las que se llegue, tengan un rango de validez lo más amplio posible. Repetir en el tiempo y/o espacio un experimento ayuda para esto, los experimentos factoriales también son útiles para este propósito.

5. **Grado de incertidumbre**: un buen experimento deber permitir calcular la probabilidad de que los resultados hayan sido obtenidos únicamente por casualidad, es decir, que debe proveer los datos suficientes y necesarios para contrastar objetivamente la hipótesis nula.

1.5 ALGUNAS RECOMENDACIONES EN LA APLICACIÓN DE LOS MÉTODOS ESTADÍSTICOS EN LA EXPERIMENTACIÓN

Montgomery (2004) cita que el uso inteligente de las técnicas estadísticas en la experimentación requiere que el investigador tenga en mente los siguientes puntos:

1. Uso del conocimiento no estadístico del problema. Generalmente los investigadores conocen a fondo su campo de especialidad. En algunos campos puede utilizarse una gran cantidad de teoría para explicar las relaciones que hay entre los factores y las respuestas. Este tipo de conocimiento no estadístico es invaluable al elegir los factores y sus niveles, al decidir el número de repeticiones que se desean realizar, al analizar los resultados, etc. La estadística no puede sustituir el hecho de reflexionar sobre el problema.

2. Mantener el diseño y el análisis tan simple como sea posible. No se debe exagerar el uso de técnicas estadísticas complejas y muy refinadas. Por lo general, lo más adecuado son los métodos de diseño y análisis relativamente simples. Si se realiza el diseño cuidadosa y correctamente, el análisis será con frecuencia, relativamente directo.

3. Reconocer la diferencia entre la significación práctica y la estadística. No hay seguridad de que una diferencia sea suficientemente grande, desde el punto de vista práctico, por el solo hecho de que dos condiciones experimentales producen respuestas medias, estadísticamente diferentes.

4. Usualmente los experimentos son iterativos. Generalmente, al inicio de un experimento no se está en condiciones de responder adecuadamente a todas las preguntas de investigación, pero es posible conocer las respuestas a medida que se avanza en la experimentación. Esto favorece el empleo del enfoque iterativo o secuencial.

1.6 VALIDEZ EXPERIMENTAL

Un experimento es válido si llega a comprobar lo que en realidad siempre ha sido cierto. En otras palabras, un experimento es válido sí:

1) Los resultados obtenidos se deben solamente a la variable independiente, y

2) Se pueden generalizar los resultados con respecto a situaciones ajenas al ambiente donde se lleva a cabo el experimento.

Cuando los resultados que se obtienen se deben solamente a la variable independiente, se dice que tienen **validez interna**. Cuando se pueden generalizar los resultados con respecto a otras situaciones se llama: **validez externa**. Una de las dificultades que se nos presentan al realizar una investigación experimental, es que es difícil (si no imposible) aumentar al máximo un tipo de validez, a la vez que el otro se aumenta al máximo.

Las variables extrañas que afectan a la variable dependiente constituyen amenazas a la validez interna de un experimento. Para poder controlar todas las variables extrañas (validez interna) sería necesario contar con una situación tan especial, que se haría difícil generalizar con respecto a otras situaciones (falta de validez externa). Si la situación experimental es muy natural y constituye un buen ejemplo de lo que son muchas situaciones (validez externa), generalmente es porque existen muchas variables extrañas (falta de validez interna).

1.6.1 Amenazas a la Validez Interna

Campbell, D. y Stanley, J. (1982) identificaron ocho variables extrañas que constituyen las principales amenazas a la validez interna de un experimento:

1) **Historia:** es todo aquello que sucede fuera de la situación experimental sobre lo cual el investigador no tiene algún control, pero que puede influir en la variable dependiente.

2) **Maduración:** se refiere al crecimiento físico, emocional e intelectual que sucede por obra natural a través del tiempo. A veces, los cambios de esta naturaleza explican que haya diferencias en la variable dependiente.

3) **Uso de tests:** constituye una amenaza a la validez interna cuando el pre-test mismo influye en la calificación que se obtenga en el post-test.

4) **Instrumentación:** se refiere a los problemas de confiabilidad que puedan existir en los instrumentos de investigación. La instrumentación puede además resultar problemática, si el post-test que se aplica es muy distinto al pre-test y si se cambia la manera de calificarlos.

5) **Regresión estadística:** puede suceder cuando los grupos se seleccionan con base en calificaciones extremas. Existe una tendencia estadística cuando se trata de grupos altos y bajos, de regresar hacia la media a través de la aplicación de otro test.

6) **Selección diferencial de sujetos:** es otro nombre que reciben algunos sesgos de muestreo. Se refiere específicamente al hecho de que si los sujetos de los grupos experimentales no se seleccionan aleatoriamente, las diferencias que se encuentren en el experimento tal vez obedezcan a diferencias iniciales que había entre los grupos y no al tratamiento experimental.

7) **Mortalidad:** es sinónimo de pérdida de individuos, es decir, otro tipo de sesgo de muestreo.

8) **Interacción:** se refiere al efecto que produce la combinación de dos o más de las otras amenazas mencionadas anteriormente.

1.6.2 Amenazas a la Validez Externa

Campbell, D. y Stanley, J. (1982) identificaron cuatro amenazas a la validez externa:

1) **Efecto reactivo del pre-test:** significa que cabe la posibilidad de que el tratamiento experimental no funcione sin haber aplicado previamente el pre-test. Si en otra situación se aplicara el tratamiento, sin el pre-test, es posible que no se obtuvieran los resultados esperados.

2) **Interacción entre la selección y el tratamiento:** es semejante a la amenaza a la validez interna que se conoce como “selección diferencial de sujetos”. La diferencia es que, quizás la diferencia entre el grupo experimental y el grupo de control sea verdadera, pero esos dos grupos son tan particulares que esa misma diferencia no va a encontrarse empleando otras muestras de la población.

3) **Situaciones reactivas:** se refieren a la manera de realizar el experimento.

4) **Interferencia de tratamientos múltiples:** se refiere a lo que puede pasar cuando los sujetos reciben más de un tratamiento. Es posible que el efecto que se haya medido después de aplicar un segundo tratamiento tenga que ver con el primer tratamiento.

En los siguientes 10 capítulos se presenta la descripción de los principales diseños experimentales utilizados en Agronomía y ciencias afines. En cada capítulo se desarrollan ejemplos detalladamente, mostrando paso a paso los cálculos realizados. Además se presentan los respectivos programas elaborados en SAS (*Statistical Analysis System*, v. 9.13) y en algunos casos salidas de Infostat v.2015 (puede consultar el manual del usuario disponible dentro del programa). Para ampliar sus conocimientos sobre el uso de *SAS*® y su aplicación en Estadística Experimental, pueden consultar las páginas web siguientes:

a) Diseño y Análisis de Experimentos usando SAS. S.P. Sinha. Instituto de Estadística Aplicada y Computación, Universidad de Los Andes, Mérida, Venezuela. http://webdelprofesor.ula.ve/economia/sinha/index.htm#beg

b) LCE-602 Estatística Experimental (aulas práticas) A.A. Franco García, D. Barbin y S. De Stefano. Departamento de Ciencias Exactas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidad de São Paulo, Brasil. http://www.lce.esalq.usp.br/sonia.html

1.7 RESEÑA HISTÓRICA Y TENDENCIAS ACTUALES EN EL ANÁLISIS DE DATOS EXPERIMENTALES

Los fundamentos de la Estadística Experimental surgieron a partir de 1919, cuando Sir Ronald Fisher (Londres, 1890-Adelaida, Australia, 1962) asumió la dirección del Departamento de Estadística de la Estación Experimental de Rothamspstead en Londres, Inglaterra. La *Rothamsted Experimental Station* fue fundada en 1843, como instituto pionero de investigación agropecuaria mundial. En 1919, el director del referido instituto decidió contratar un matemático para crear un departamento de Estadística, con el objetivo de analizar la gran cantidad de datos acumulados de los llamados "experimentos clásicos" instalados desde 1843. Fisher fue contratado y permaneció como jefe de este departamento durante un largo período. Estando en Rothamsted, escribió dos artículos de gran importancia, uno en 1922 sobre máxima verosimilitud, y otro en 1925 sobre pequeñas muestras experimentales, aclarando la diferencia entre estadística muestral y valores de población.

En 1925 Fisher desarrolló el análisis de varianza (ANOVA) con implicaciones en la estimación de los componentes de varianza y diseños experimentales. En 1926 enfantizó el papel crucial de la repetición, aleatorización y del control local en la eficiencia de los experimentos. Creó entonces el diseño de bloques completos al azar.

Yates (1940) creó los diseños en bloques incompletos o látices, los cuales permiten un mejor control de la heterogeneidad experimental, que es equivalente al análisis de modelos mixtos con bloques aleatorios. Eisenhart (1947) identificó formalmente efecto aleatorio, efecto fijo y modelos mixtos.
Siendo así, que los métodos experimentales surgieron primero en las ciencias biológicas y a partir de entonces se generalizaron para todas las áreas del conocimiento humano. A mediados de la década de 1950, se tenía prácticamente definido el conjunto de esquemas experimentales, con los cuales se podía resolver cualquier problema de la investigación agronómica.

Henderson (1953) fue el primero en usar explícitamente la metodología de los modelos mixtos para estudios de mejoramiento genético animal. Harville (1976, 1977) publicó la teoría formal y completa de los modelos mixtos. Aunque el análisis de casos especiales de respuestas no normalmente distribuidas como análisis probit (Bliss, 1935) y análisis Logit (Berkson, 1944) existían en el contexto de bioensayos, textos sobre métodos estadísticos estándar, tales como Steel et al. (1997) y Snedecor y Cochran (1989) trataban el problema de falta de normalidad por medio del uso de transformaciones. El propósito de las transformaciones, tal como logarítmica, arcoseno y raíz cuadrada fue permitir al investigador obtener análisis aproximados usando los métodos estándar de la teoría normal. Box y Cox (1964) propusieron una clase general de trasnformaciones que incluia las antes mencionadas como casos especiales. Con la misma finalidad, de permitir el uso de los métodos estadísticos basados en la distribución normal.

Nelder y Wedderburn (1972) articularon la teoría completa de modelos lineales para variables de respuesta que no siguen una distribución normal. Ellos asumieron que la distribución de probabilidad de la variable de respuesta pertenece a la familia exponencial. Esta familia de distribuciones de probabilidad contiene un diverso conjunto de distribuciones continuas (normal, lognormal, gamma, beta) y discretas (binomial, multinomial, Poisson, binomial negativa). Estos modelos fueron referidos como Modelos Lineales Generalizados –MLG– (no debe ser confundido este término con el de Modelos Lineales Generales, que es utilizado para variables de respuesta que siguen solamente una distribución normal). McCullagh y Nelder (1989) dieron inicio a la investigación de los modelos lineales generalizados de efectos mixtos. El *software* SAS tiene incorporados los procedimientos: MIXED (para modelos mixtos), GENMOD (para MLG con efectos fixos) y GLIMMIX (para MLG con efectos mixtos). En los programas R e Infostat también es posible trabajar estas metodologías.

CAPÍTULO 2
DISEÑO COMPLETAMENTE AL AZAR

2.1 INTRODUCCIÓN

En este tipo de diseño están incluidos los principios de repetición y de aleatorización, o sea que, es utilizado cuando no hay necesidad del control local, debido a que el ambiente experimental y las condiciones de manejo son homogéneos y los tratamientos se asignan a las unidades experimentales mediante una aleatorización completa, sin ninguna restricción.

2.1.1 Ventajas

a. La estructura del análisis estadístico es simple.
b. Permite máxima flexibilidad en cuanto al número de tratamientos y número de repeticiones.
c. La pérdida de observaciones durante la conducción del experimento no genera dificultades en el análisis y en la interpretación de los resultados.
d. Reúne el mayor número de grados de libertad en el residuo, en comparación con otros diseños.

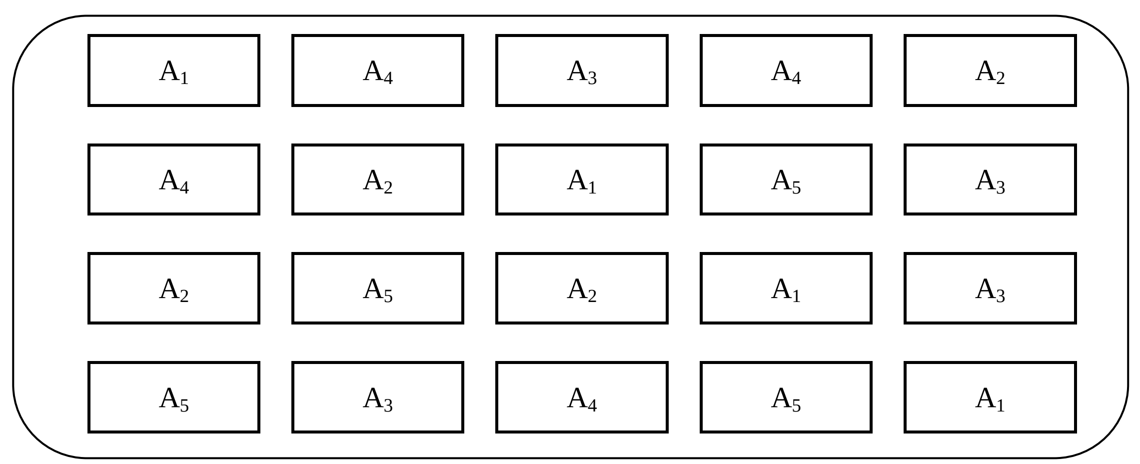
2.1.2 Inconvenientes

a. Cuando el número de unidades experimentales es muy grande es difícil encontrar lugares grandes que presenten la homogeneidad requerida.

b. Debido a que las fuentes de variación no asociadas a los tratamientos o a los niveles del factor en estudio, están incluidas en el residuo como variación del azar, la buena precisión de los análisis se ve comprometida.

2.1.3 Aleatorización

Considerando un experimento con t = 5 niveles del factor A (tratamientos) y r = 4 repeticiones para cada nivel, se tiene que el número total de unidades experimentales (parcelas) incluidas en el experimento es t × r = 5 × 4 = 20. Las (t × r) parcelas serán aleatorizadas sin restricciones, los t niveles del factor A en estudio con sus r repeticiones, conforme se muestra en el siguiente croquis.



Las respuestas obtenidas en función de la aplicación de cada nivel del factor A en estudio en sus respectivas repeticiones pueden ser representadas por y_{ij}, que es considerada como una variable aleatoria.

2.2 ANÁLISIS ESTADÍSTICO

A continuación se muestra la representación de las observaciones de un experimento, con un factor con *t* tratamientos (o niveles) y *r* repeticiones.

Tratamientos	Repeticiones					$y_{i.}$
	1	2	3	. . .	*r*	
1	y_{11}	y_{12}	y_{13}	. . .	y_{1r}	$y_{1.}$
2	y_{21}	y_{22}	y_{23}	. . .	y_{2r}	$y_{2.}$
3	y_{31}	Y_{32}	y_{33}	. . .	y_{3r}	$y_{3.}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
t	y_{t1}	y_{t2}	y_{t3}	. . .	y_{tr}	$y_{t.}$

2.2.1 Hipótesis

$$\bar{Y}_{..} = \frac{Y_{..}}{tr}$$

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)

Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (al menos uno de los tratamientos produce efectos distintos)

2.2.2 Modelo Estadístico

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \begin{cases} i = 1,2, \ldots t \\ j = 1,2, \ldots r \end{cases}$$

siendo,

Y_{ij} = variable de respuesta de la ij-ésima unidad experimental

μ = media general de la variable de respuesta

τ_i = efecto del i - ésimo tratamiento (nivel del factor) en la variable dependiente.

ε_{ij} = error experimental asociado a la ij-ésima unidad experimental

2.2.3 Supuestos

Las suposiciones que validan el análisis de varianza son:

a. Los errores son independientes.
b. Los errores están normalmente distribuidos con media cero y varianza constante
c. Existe homogeneidad de varianzas entre los tratamientos
d. El modelo es lineal y de efectos aditivos.

2.2.4 Descomposición de la suma de cuadrados total

El análisis de varianza es un proceso aritmético y estadístico, que consiste en descomponer la variación total en fuentes o causas de variación. Por variación total se entiende, la variación entre las unidades experimentales (o parcelas).

La variabilidad total de las observaciones Y_{ij}, cuando no se considera la información acerca de los tratamientos, es medida en términos de la desviación total de cada observación, esto es, la desviación de los Y_{ij} alrededor de la media general $\bar{Y}_{..}$:

$$Y_{ij} - \bar{Y}_{..} \qquad \textbf{(1)}$$

Cuando se utiliza información acerca de tratamientos, las desviaciones entre cada observación Y_{ij} alrededor de la media estimada de su respectivo tratamiento, reflejan la incertidumbre restante en los datos, y es dada por:

$$Y_{ij} - \bar{Y}_{i.} \qquad \textbf{(2)}$$

La diferencia entre las desviaciones (1) y (2) refleja la diferencia entre la media estimada de tratamientos y la media general:

$$\left(Y_{ij} - \bar{Y}_{..}\right) - \left(Y_{ij} - \bar{Y}_{i.}\right) = \bar{Y}_{i.} - \bar{Y}_{..} \qquad \textbf{(3)}$$

Note que a partir de la ecuación (3), se puede descomponer la desviación total $Y_{ij} - \bar{Y}_{..}$ en dos componentes:

$$Y_{ij} - \bar{Y}_{..} \quad = \quad \bar{Y}_{i.} - \bar{Y}_{..} \quad + \quad Y_{ij} - \bar{Y}_{i.} \qquad \textbf{(4)}$$

Desviación total	Desviación de la media estimada de tratamiento alrededor de la media general	Desviación alrededor de la media estimada de tratamiento.

A partir de esto, la desviación total $Y_{ij} - \bar{Y}_{..}$ puede ser vista como la suma de dos componentes:

1. La desviación de la media estimada de tratamientos alrededor de la media general.

2. La desviación de Y_{ij} alrededor de la media estimada de su tratamiento, que es simplemente, el residuo e_{ij}.

A partir de la ecuación (4) se pueden obtener las expresiones matemáticas utilizadas para calcular las sumas de cuadrados de tratamientos, error experimental y total. Para realizarlo, se inicia obteniendo la suma de cuadrados de la desviación total:

$$(Y_{ij} - \bar{Y}_{..})^2 = [(\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})]^2$$

Luego, se obtiene la sumatoria de las desviaciones totales al cuadrado:

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}\left[(\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})\right]^2 \quad \textbf{(5)}$$

Como $\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..})^2$ = Suma cuadrados total, y desarrollando el binomio de lado derecho de la ecuación (5), ésta quedaría así:

$$SC_{total} = \sum_{i=1}^{t}\sum_{j=1}^{r}\left[(\bar{Y}_{i.} - \bar{Y}_{..})^2 + 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.}) + (Y_{ij} - \bar{Y}_{i.})^2\right] \quad \textbf{(6)}$$

Sumando primero sobre j, la ecuación (6) queda:

$$SC_{total} = \sum_{i=1}^{t}\left[r\,(\bar{Y}_{i.} - \bar{Y}_{..})^2 + 2\,(\bar{Y}_{i.} - \bar{Y}_{..})\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.}) + \sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^2\right]$$

Donde: $\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.}) = \sum_{j=1}^{r}Y_{ij} - r\bar{Y}_{i.} = Y_{i.} - r\frac{Y_{i.}}{r} = Y_{i.} - Y_{i.} = 0$, por tanto

$2\,(\bar{Y}_{i.} - \bar{Y}_{..})\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})$ = 0, entonces:

$$SC_{total} = r\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^2$$

Suma de cuadrados de tratamientos — Suma de cuadrados del error experimental

$$SC_{trat} = r\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$= r\sum_{i=1}^{t}(\bar{Y}_{i.}^2 - 2\bar{Y}_{i.}\bar{Y}_{..} + \bar{Y}_{..}^2) = r\sum_{i=1}^{t}\bar{Y}_{i.}^2 - 2r\bar{Y}_{..}\sum_{i=1}^{t}\bar{Y}_{i.} + (tr)\bar{Y}_{..}^2 \quad \textbf{(7)}$$

Para el primer término de la ecuación (7), como: $\overline{\overline{Y}}_{i.} = \frac{Y_i.}{r}$, $r\sum_{i=1}^{t}\overline{\overline{Y}}_{i.}^{\,2} = r\sum_{i=1}^{t}\frac{Y_i.^2}{r^2} = \sum_{i=1}^{t}\frac{Y_i.^2}{r}$

En el caso del segundo término $-2r\,\overline{\overline{Y}}..\sum_{i=1}^{t}\overline{\overline{Y}}_{i.}$, se sabe que $\overline{\overline{Y}}_{i.} = \sum_{i=1}^{t}\frac{Y_{ij}}{r}$, se tiene que:

$\sum_{i=1}^{t}\overline{\overline{Y}}_{i.} = \sum_{i=1}^{t}\sum_{j=1}^{r}\frac{Y_{ij}}{r}$, y $\overline{\overline{Y}}.. = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{tr}$ entonces:

$$-2r\,\overline{\overline{Y}}..\sum_{i=1}^{t}\overline{\overline{Y}}_{i.} = -2r\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{tr}\sum_{i=1}^{t}\sum_{j=1}^{r}\frac{Y_{ij}}{r} = -2\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{tr}\,\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{} =$$

$$-2\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{tr}\,\frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{} = -2\frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr}$$

Para finalizar, analicemos el tercer término:

Sí $\overline{\overline{Y}}.. = \frac{Y..}{tr} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}}{tr}$, entonces $(tr)\overline{\overline{Y}}..^2 = (tr)\ \frac{Y..^2}{(tr)^2} = \frac{Y..^2}{(tr)} = \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr}$, entonces:

$$-2\frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr} + \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr} = -\frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr}\ .$$

La expresión de suma de cuadrados para tratamientos (o sea, entre tratamientos) queda de la siguiente manera:

$$\sum_{i=1}^{t}\frac{Y_i.^2}{r} - \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}\right)^2}{tr}.$$

Para el caso de la suma de cuadrados total: $\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij}-\overline{\overline{Y}}..)^2$, se tiene al desarrollar el binomio:

$SC_{total} = \sum_{i=1}^{t}\sum_{j=1}^{r}\left[(Y_{ij})^2 - 2(Y_{ij}\ \overline{\overline{Y}}..) + (\overline{\overline{Y}}..)^2\right]$, y al distribuir las sumatorias:

$$SC_{total} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}^2 - 2\overline{Y}..\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij} + tr\,\overline{Y}..^2 ,$$ y como:

$$(tr)\overline{Y}..^2 = (tr)\frac{Y..^2}{(tr)^2} = \frac{Y..^2}{(tr)} = \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr}$$, la suma de cuadrados total queda:

$$SC_{total} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}^2 - 2\frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr} + \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}^2 - \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr}$$

La suma de cuadrados del error (o sea, dentro de tratamientos), se obtiene por diferencia: $SC_{ee} = SC_{total} - SC_{trat.}$,

$$SC_{ee} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}^2 - \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr} - \sum_{i=1}^{t}\frac{Y_{i}.^2}{r} + \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}\right)^2}{tr} .$$

$$SC_{ee} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}^2 - \sum_{i=1}^{t}\frac{Y_{i}.^2}{r}$$

2.2.5 Prueba de F

La estadística F es definida como la razón de dos variables aleatorias independientes con distribución χ^2 (Ji-cuadrada o *Chi-Square*), cada una de ellas dividida por sus respectivos grados de libertad, o sea:

$$F = \frac{Q_1 / n_1}{Q_2 / n_2} ,$$

siendo Q_1 una variable aleatoria con distribución χ^2 y n_1 grados de libertad y Q_2 una variable aleatoria con distribución χ^2 y n_2 grados de libertad, ambas independientes.

Así, considerando que SC_{trat}/σ^2 tiene distribución χ^2 con (t–1) grados de libertad, bajo H_o, y SC_{eet}/σ^2 tiene distribución χ^2 con t (r–1) grados de libertad y, son independientes, entonces el cociente entre esas variables aleatorias, o sea:

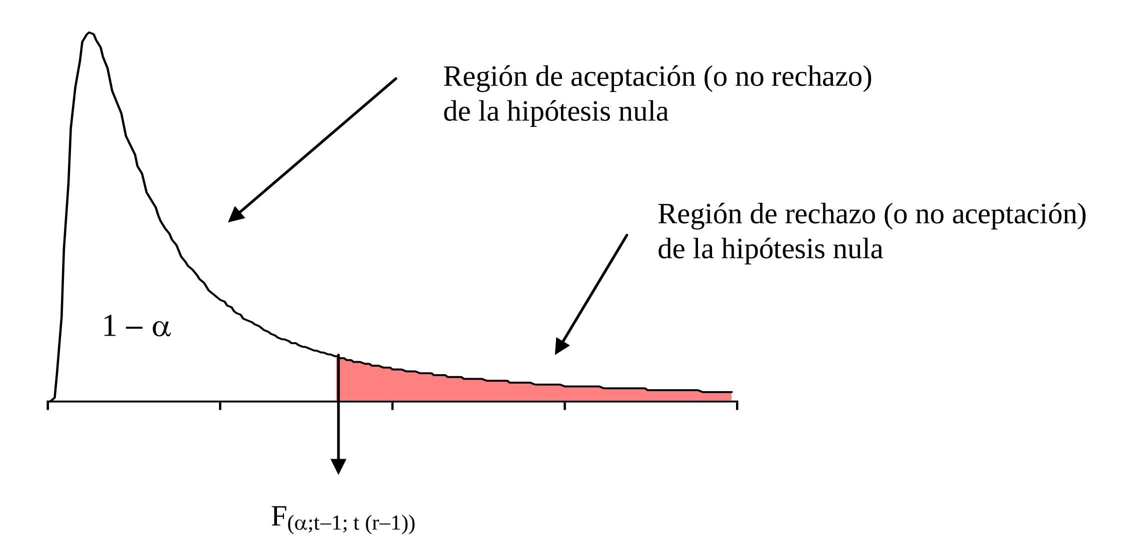
$$\frac{SC_{trat} /(t-1)}{SC_{ee} / t(r-1)} = \frac{CM_{trat}}{CM_{ee}} = F_o ,$$

tiene distribución F de Fisher & Snedecor con (t–1) y t (r–1) grados de libertad, bajo H_o. Ese cociente es la estadística apropiada para evaluar la H_o, o sea: $t_1 = . \ . \ . = t_I = 0$.

Así, se puede decidir por el rechazo de H_o, al nivel α de significancia sí:

$$F_o \geq F_{(\alpha;\, t-1;\, t\,(r-1))},$$

en que $F_{(\alpha;t-1;\ t\ (r-1))}$, es el cuantil de orden $(1-\alpha)$ de la distribución F con $(t-1)$ y $t\ (r-1)$ grados de libertad, como se muestra en la gráfica siguiente:



Sí el valor observado de F (F_o) es superior al valor crítico $F_{(\alpha;t-1;\ t\ (r-1))}$, la H_o es rechazada, por lo tanto se concluye que existen diferencias significativas entre los efectos de los tratamientos.

2.2.6 Análisis de varianza o variación con aplicación de la prueba de F

El esquema del análisis de varianza abreviado como ANDEVA, ANOVA (**AN**alysis **O**f **VA**riance) o bien ANVA; y las expresiones necesarias para la aplicación de la estadística F, para la prueba de hipótesis se presentan en el siguiente cuadro.

Fuentes de variación (FV)	Grados de libertad (gl)	Suma de Cuadrados (SC)	Cuadrados Medios (CM)	Valor de F
Tratamientos	t –1	$\sum_{i=1}^{t} \frac{Y_{i.}^2}{r} - \frac{Y_{..}^2}{tr}$	SC_{trat} / gl $_{trat}$	CM_{trat} / CM_{ee}
Error	t (r - 1)	$SC_{total} - SC_{trat}$	SC_{ee} / gl $_{ee}$	
Total	t r - 1	$\sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij}^2 - \frac{Y_{..}^2}{tr}$		

Regla de Decisión

Rechazar Ho. Si Valor de F $\geq$ F crítico $_{(gl\ trat;\ gl\ error;\ \alpha)}$

No Rechazar Ho. Si Valor de F $<$ F crítico $_{(gl\ trat;\ gl\ error;\ \alpha)}$

$F_{crítica}$ = Valor crítico de F encontrado en la tabla F de Fisher & Snedecor, considerando los grados de libertad de tratamientos (v_1), los grados de libertad del error (v_2) y un determinado nivel de significancia (α)

2.2.7 Coeficiente de Variación (CV)

Se le puede considerar como medida relativa de la variación que no es posible controlar en el experimento (error experimental) y se calcula de la siguiente forma:

$$CV = \frac{\sqrt{CM_{ee}}}{\bar{Y}_{..}} \times 100$$

El coeficiente de variación da una idea de la precisión del experimento, a un valor alto de CV corresponde un alto error experimental, lo cual indica que existe poca capacidad del experimento para detectar diferencias significativas entre los tratamientos.

"De modo general, altos coeficientes de variación indican experimentos mal manejados", pero no siempre. El hecho de que el coeficiente de variación sea alto puede deberse no solamente al mal manejo del experimento, sino también a:

a) tipo de variable de respuesta (escala de medición),
b) tipo de tratamientos,
c) errores en el análisis de la información, etc.

El CV puede ser utilizado para comparar la precisión experimental de variables experimentales semejantes. Es conveniente que el investigador revise bibliografía sobre los valores de coeficiente de variación obtenidos en cada cultivo y condición donde se realizó el experimento (por ejemplo, un valor de CV=10% puede ser considerado un valor bajo, pero para algunas condiciones no).

Considere la siguiente situación donde se tienen las siguientes observaciones de una variable de respuesta:

	Variable de respuesta	Variable de respuesta transformada (con raíz cuadrada)
	36	6
	16	4
	9	3
	4	2
Media	16.25	3.75
Desviación estándar	14.06	1.71
CV%	86.50	45.54

Observe que hubo una reducción del valor del CV% para la variable transformada. Por eso, el CV% no siempre es un buen indicador de la precisión experimental (Dos Anjos, 2003). Para más información sobre control de calidad de experimentos puede consultar el texto de Storck *et al.* (2011).

2.2.8 Ejemplo de Aplicación

Un silvicultor quiso comparar los efectos de cinco tratamientos de preparación del terreno sobre el crecimiento inicial en altura de plántulas de pino *maximinoii*. Dispuso de 25 parcelas y aplicó cada tratamiento a cinco parcelas seleccionadas al azar. La plantación fue realizada manualmente y, al final de cinco años, se midió la altura de todos los pinos y se calculó la altura promedio de cada parcela. Las medidas de las parcelas (en pies) fueron como sigue:

	Tratamientos					
	A	B	C	D	E	
	15	16	13	11	14	
	14	14	12	13	12	
	12	13	11	10	12	
	13	15	12	12	10	
	13	14	10	11	11	
$Y_{i.}$	67	72	58	57	59	313
$\bar{Y}_{i.}$	13.4	14.4	11.6	11.4	11.8	12.52

$$SC_{trat} = \frac{67^2 + 72^2 + 58^2 + 57^2 + 59^2}{5} - \frac{313^2}{25} = 34.64$$

$$SC_{total} = 15^2 + 14^2 + \ldots + 11^2 - \frac{313^2}{25} = 64.24$$

$$SC_{error} = 64.24 - 34.64$$

El resumen del análisis de varianza es presentado a continuación:

Fuentes de variación	**Grados de libertad**	**Suma de cuadrados**	**Cuadrados Medios**	**F**	**Valor crítico F**
Tratamientos	4	34.64	8.66	5.85	2.87
Error Experimental	20	29.6	1.48		
Total	24	64.24			

CV = 9.72%

Nota: El valor crítico de F lo puede obtener directamente de la Tabla 1 del Apéndice, o en MS Excel®, en el menú INSERTAR busque función (***fx***) y luego seleccione la categoría **Estadísticas**, y dentro de éstas DISTR.F.INV (probabilidad, grados_de_libertad1, grados_de_libertad2). Otra manera de poder concluir, es obteniendo el valor p (*p value*), si éste es menor a 0.05, se rechaza la hipótesis nula. Para este caso valor p = 0.0027607. Más información sobre el valor p y su interpretación puede encontrarla en el texto del curso de Estadística Geral (LÓPEZ BAUTISTA, 2010).

Conclusión:

Los tratamientos de preparación en el sitio afectan significativamente el crecimiento inicial en altura de plántulas de pino en el terreno; debido a que el valor de F es superior al valor crítico. Se recomienda realizar un análisis posterior al ANOVA para poder identificar el mejor tratamiento.

2.3 ANÁLISIS DE VARIANZA PARA UN EXPERIMENTO REALIZADO EN UN DISEÑO COMPLETAMENTE AL AZAR DESBALANCEADO

El diseño completamente al azar es "NO BALANCEADO" (o DESBALANCEADO), cuando los niveles del factor en estudio no poseen el mismo número de repeticiones, debido a parcelas perdidas o a la falta de material experimental.

En este modelo, el hecho de no ser balanceado no trae alteraciones en el proceso del ANOVA, pero las pruebas utilizadas en las comparaciones múltiples pasan a ser apenas aproximadas.

Modelo estadístico: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ $\begin{cases} i = 1, 2, \ldots, t \\ j = 1, 2, \ldots, r_i \end{cases}$

siendo r_i el número de repeticiones del tratamiento i, y $\sum_{i=1}^{t} r_i = n$, el total de unidades experimentales involucradas en el experimento.

Las hipótesis y los supuestos no varían, con respecto al DCA balanceado.

CUADRO DE RESUMEN DEL ANÁLISIS DE VARIANZA

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados Medios	F
Tratamientos	$t-1$	$\sum_{i=1}^{t} \frac{Y_{i.}^2}{r_i} - \frac{Y_{..}^2}{n}$	SC_{trat} / gl_{trat}	CM_{trat} / CM_{ee}
Error Experimental	$\sum_{i=1}^{t} r_i - t$	$SC_{total} - SC_{trat}$	SC_{ee} / gl_{ee}	
Total	$\sum_{i=1}^{t} r_i - 1$	$\sum_{i=1}^{t} \sum_{j=1}^{r_i} Y_{ij}^2 - \frac{Y_{..}^2}{n}$		

$$Y_{..} \sum_{i=1}^{t} \sum_{j=1}^{r_i} Y_{ij} \qquad \bar{Y}_{..} = \frac{\sum_{i=1}^{t} \sum_{j=1}^{r_i} Y_{ij}}{\sum_{i=1}^{t} r_i} \qquad \bar{Y}_{i.} = \frac{\sum_{j=1}^{r_i} Y_{ij}}{r_i}$$

Ejemplo de Aplicación

Considere las siguientes producciones diarias (kg) de leche (con 4% de grasa) de vacas para lactación, sometidas a la administración de raíces y tubérculos, como suplemento de invierno en la alimentación; datos de un experimento citado por Gomes, FP (2000).

	SIN SUPLEMENTO	MANDIOCA (*Manihot esculenta*)	ARARUTA (*Maranta arundinacea*)	BATATA DOCE (*Ipomoea batata*)
	19.58	23.40	35.43	22.15
	21.07	22.37	32.47	24.37
	23.43	24.36	34.48	26.54
	25.42	25.12	33.79	20.37
	22.81	22.94	35.04	19.54
	23.52		35.19	24.06
$Y_{i..}$	135.83	118.19	206.40	137.03
$\bar{Y}_{i.}$	22.64	23.64	34.40	22.84
R	6	5	6	6

$$SC_{trat} = \left(\frac{135.83^2}{6} + \frac{118.19^2}{5} + \frac{206.40^2}{6} + \frac{137.03^2}{6}\right) - \frac{597.45^2}{23} = 579.02$$

$$SC_{total} = \left(19.58^2 + 21.07^2 + \ldots + 24.06^2\right) - \frac{597.45^2}{23} = 646.06$$

$$SC_{ee} = 646.06 - 579.02 = 67.04$$

$$\bar{Y}_{..} = \frac{597.45}{23} = 25.98 \text{ kg}$$

RESUMEN DEL ANÁLISIS DE VARIANZA

Fuentes de variación	**Grados de libertad**	**Suma de cuadrados**	**Cuadrados Medios**	**Valor de F**	**F(3, 19, 0.01)**
Suplemento	3	579.02	193.01	54.68**	5.01
Error experimental	19	67.04	3.53		
Total	22	646.06			

** significativo al 1% CV = 7.23%

Conclusión: Con un nivel de significancia de 1% se rechaza la H_o, verificándose que existe efecto de la suplementación alimentícia sobre la producción diaria de leche.

2.4 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR

```
OPTIONS nodate nonumber; /*para que en la salida no aparezca fecha ni paginación*/
DATA dca;
INPUT trat $ alt; /*el signo de dólar $ indica que la variable es de tipo alfanumérica*/
LABEL alt = "altura del árbol en pies"; /*Label indica la etiqueta, el nombre de la variable alt*/
CARDS;
A	15
A	14
A	12
A	13
A	13
B	16
B	14
B	13
B	15
B	14
C	13
C	12
C	11
C	12
C	10
D	11
D	13
D	10
D	12
D	11
E	14
E	12
E	12
E	10
E	11
;
PROC anova;
CLASS trat;
MODEL alt = trat;
MEANS trat/Tukey;
RUN;
```

DATA: es el nombre del archivo temporal que utilizará SAS para ejecutar el programa.

INPUT: aquí se enlistan las variables que incluye la base de datos a analizar. En este caso una columna corresponde a trat (tratamientos) y otra a alt (altura). El signo $ sirve para indicar que la variable trat es alfanumérica.

CARDS: indica que a continuación se presentan los datos.

PROC: abreviatura de PROCEDURE, indica el tipo de procedimiento solicitado a SAS, en este caso ANOVA (análisis de varianza)

CLASS: aquí se indica cuáles son las variables independientes.

MODEL: describe el modelo a analizar.

MEANS: esta instrucción se utiliza para indicar el tipo de prueba de comparación múltiple de medias se va a solicitar a SAS que ejecute. Además de Tukey, otras opciones pueden ser: BON, DUNCAN, LSD, SCHEFFE, SNK o T.

En el caso de datos faltantes la opción es: **MEANS trat/tukey lines;**
La opción *lines* genera la media armónica de las repeticiones para ser usada en la prueba de Tukey.

En el capítulo 3 se ampliará el tema de las pruebas de comparación múltiple de medias.

Observación: En el caso de presentarse datos faltantes, se pueden presentar dos situaciones:

1. Ausencia de datos: por ejemplo, sí la variable de respuesta fue número de insectos vivos (o muertos) por planta y al momento de realizar el muestreo no se encontró presencia del insecto; el valor reportado será 0.

2. Parcela perdida: retomando el ejemplo citado en el inciso anterior, en caso de que, al realizar la toma de datos, exista pérdida de las plantas, en la base de datos se digitará un punto (.) para indicarle al programa que se trata de una observación perdida.

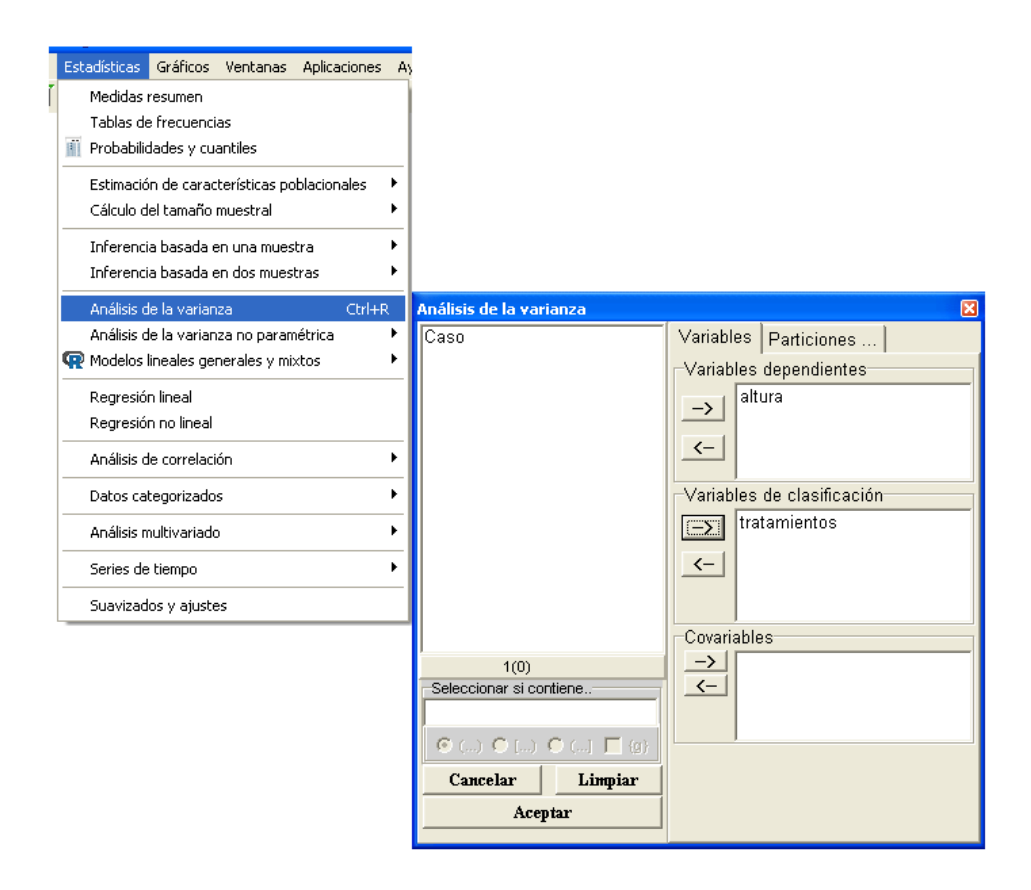
2.5 INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR

Los datos del ejercicio 2.2.8 quedarían organizados de la siguiente manera en el entorno de Infostat:

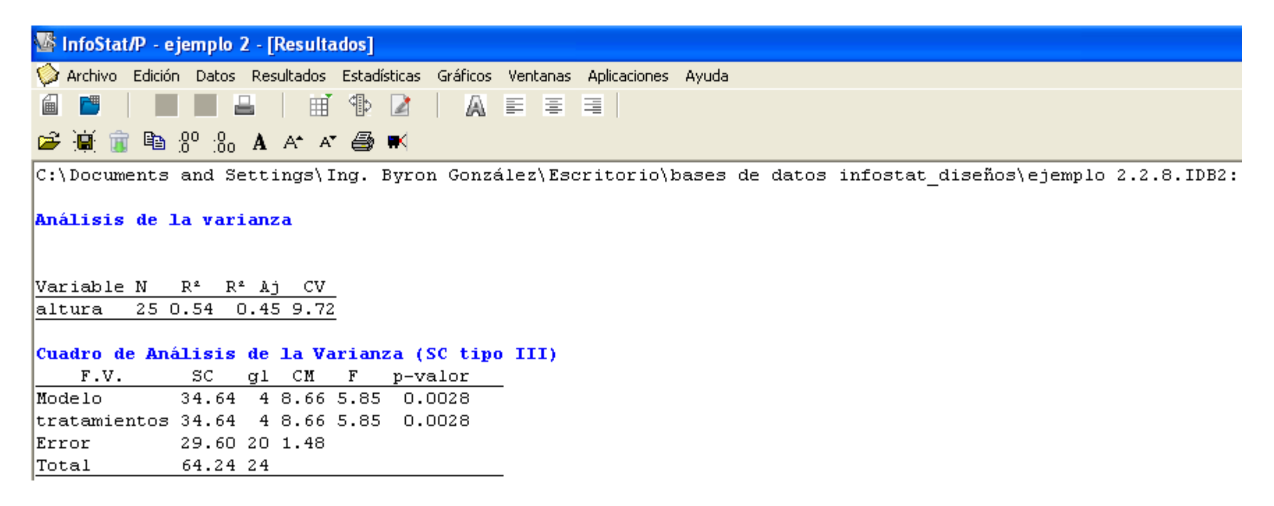
Archivo Edición Datos R

Caso	tratamientos	altura
1	A	15
2	B	16
3	C	13
4	D	11
5	E	14
6	A	14
7	B	14
8	C	12
9	D	13
10	E	12
11	A	12
12	B	13
13	C	11
14	D	10
15	E	12
16	A	13
17	B	15
18	C	12
19	D	12
20	E	10
21	A	13
22	B	14
23	C	10
24	D	11
25	E	11

Para solicitar los resultados del análisis de varianza se sigue la ruta:



Y los resultados obtenidos son:



C:\Documents and Settings\Ing. Byron González\Escritorio\bases de datos infostat_diseños\ejemplo 2.2.8.IDB2:

Análisis de la varianza

Variable	N	R²	R² Aj	CV
altura	25	0.54	0.45	9.72

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	34.64	4	8.66	5.85	0.0028
tratamientos	34.64	4	8.66	5.85	0.0028
Error	29.60	20	1.48		
Total	64.24	24			

Compare los resultados con los obtenidos de forma manual.

2.6 ANÁLISIS DE VARIANZA PARA UN EXPERIMENTO REALIZADO MEDIANTE UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO

En la experimentación agronómica en algunas situaciones no es posible realizar la medición de todos los individuos que conforman la unidad experimental, debido al tamaño de la misma o al costo de las mediciones. Por lo que se hace necesario tomar muestras de elementos que existen en cada parcela o unidad experimental. De esta manera tendremos más de un dato por unidad experimental, sin que esto constituya una repetición.

En este caso, el proceso para obtener estos datos se denomina: **muestreo**. Considerando el caso más simétrico y tal vez el más útil, cuando se tienen *r* repeticiones para cada uno de los *t* tratamientos y se toman *m* muestras dentro de cada unidad experimental, se tendrán en total *trm* observaciones. Algunos ejemplos de muestreo se presentan a continuación:

a) En un experimento de campo, el investigador puede no tener tiempo para cosechar (totalmente) cada unidad experimental. De esta manera, podrá seleccionar al azar varios cuadros por parcela y cosechar el grano en cada cuadro seleccionado. De nuevo, describiríamos estas observaciones como "muestras dentro de unidades experimentales".

b) En un experimento de tecnología de alimentos que implicó el almacenamiento de fresas congeladas, se almacenaron 10 cajas (unidades experimentales) a cada cinco lapsos de almacenamiento (tratamientos). Cuando se hicieron las determinaciones del ácido ascórbico después del almacenamiento, se hicieron dos determinaciones en caja (muestras dentro de unidades experimentales).

El modelo lineal apropiado para interpretar los resultados de un experimento como los descritos es:

1. Modelo Estadístico: $Y_{ijk} = \mu + \tau_i + \varepsilon_{j(i)} + \eta_{k(ij)}$ $\begin{cases} i = 1, 2, \ldots, t \\ j = 1, 2, \ldots, r \\ k = 1, 2, \ldots, m \end{cases}$

En que:

Y_{ijk} = valor de la variable de respuesta correspondiente a la k-ésima muestra sobre la unidad experimental que lleva el tratamiento i en la repetición j.
μ = Media general de la variable respuesta.
τ_i = Efecto del i-ésimo tratamiento.
$\varepsilon_{j(i)}$ = error experimental asociado a la ij-ésima unidad experimental (error entre parcelas)
$\eta_{k(ij)}$ = error de muestreo dentro de la ij-ésima unidad experimental (error dentro de parcelas).

2. Hipótesis

Ho: $\vee_i$ i = 1,2,. . . / $\tau_i = \tau$
Ha: $\exists_i$ i = 1,2,. . . / $\tau_i \neq \tau$

3. Supuestos:

a) $\varepsilon_{ij} \sim NID\ (0, \sigma_e^2)$
ε_{ij} son los errores de parcela, y se asume que son variables aleatorias no correlacionadas, con media 0 y varianza constante desconocida σ_e^2 (varianza residual o varianza entre parcelas)

b) $\eta_{ijk} \sim NID\ (0, \sigma_m^2)$

η_{ijk} son los errores de muestreo, y se consideran como variables aleatorias no correlacionadas entre sí, ni con los errores de parcela, con media 0 y varianza constante desconocida σ_m^2 (varianza muestral o varianza dentro de parcelas)

4. Análisis de varianza

Fuentes de variación	**Grados de libertad**	**Suma de cuadrados**	**Cuadrados Medios**	**F**
Tratamientos	t –1	$\sum_{i=1}^{t} \frac{Y_{i..}^2}{rm} - \frac{Y...^2}{rtm}$	SC_{trat} / gl $_{trat}$	F_2
Error Experimental	t (r–1)	$\sum_{i=1}^{t}\sum_{j=1}^{r} \frac{Y_{ij.}^2}{m} - \sum_{i=1}^{t} \frac{Y_{i..}^2}{rm}$	SC_{ee} / gl $_{ee}$	$F_1 = \frac{CM_{ee}}{CM_{em}}$
Error de Muestreo	tr (m–1)	$\sum_{i=1}^{t}\sum_{j=1}^{r}\sum_{k=1}^{m} Y_{ijk}^2 - \sum_{i=1}^{t}\sum_{j=1}^{r} \frac{Y_{ij.}^2}{m}$	SC_{em} / gl $_{em}$	
Total	trm – 1	$\sum_{i=1}^{t}\sum_{j=1}^{r}\sum_{k=1}^{m} Y_{ijk}^2 - \frac{Y...^2}{rtm}$		

$$Y... = \sum_{i=1}^{t}\sum_{j=1}^{r}\sum_{k=1}^{m} Y_{ijk} \qquad \bar{Y}... = \frac{Y...}{trm}$$

1° Pruebas preliminares de significancia

Para evaluar el efecto de muestreo, se realiza una prueba de hipótesis:

Ho: $\sigma_e^2 = 0$
Ha: $\sigma_e^2 > 0$,

comparando el cuadrado medio del error experimental, contra el cuadrado medio del error de muestreo. Bajo la hipótesis nula, el cociente F_1 = (CMee / CMem), se distribuye como una F con t(r-1) y tr(m-1) grados de libertad. Luego se presentan dos alternativas:

- Si F1 ≥ F (glee, glem, α), se rechaza Ho, indicando que el muestreo fue efectivo, en otras palabras, la varianza entre plantas dentro de parcelas es mayor que la varianza entre parcelas, por lo que F_2 se calcula de la siguiente forma: $F_2 = \frac{CM_{trat}}{CM_{ee}}$, y se compara con el valor crítico de F(glt, glee, α). El coeficiente de variación se obtiene así: $CV = \frac{\sqrt{CMee}}{\bar{Y}...} \times 100$

- Si F_1 < F (glee, glem, α), se acepta Ho, lo cual indica que el submuestreo no fue efectivo o no es importante en este experimento, por lo que los errores deben mancomunarse así:

$$CMep = \frac{SCee + SCem}{glee + glem},$$

y F_2 se obtiene de la siguiente manera: $F_2 = \frac{CM_{trat}}{CM_{ep}}$, y se compara con: $F_{crítica}$ (glt, glep, α).

El coeficiente de variación se obtiene de la siguiente manera:

$$CV = \frac{\sqrt{CMep}}{\bar{Y}...} \times 100$$

2.6.1 Ejemplo de Aplicación

Considere los siguientes resultados (datos hipotéticos), obtenidos en un experimento con tres tratamientos (A, B y C) con 4 repeticiones y 2 muestras, donde fue utilizado un DCA con muestreo:

Tratamientos	**Repeticiones**					
	I	**II**	**III**	**IV**	$Y_{i..}$	$\bar{Y}_{i..}$
A	5.6 5.7	5.0 5.1	5.5 5.4	5.3 5.5		
$Y_{ij.}$	**11.30**	**10.1**	**10.9**	**10.8**	43.10	5.39
B	6.7 8.7	4.7 3.7	5.7 6.5	6.2 5.8		
$Y_{ij.}$	**15.4**	**8.4**	**12.2**	**12.0**	48	6.00
C	7.6 7.8	7.4 7.2	7.5 7.6	5.7 6.7		
$Y_{ij.}$	**15.4**	**14.6**	**15.1**	**12.40**	57.5	7.19
					148.6	6.19

t = 3 (A, B, C)
r = 4 (I,II,III,IV)
m= 2 (submuestras)

Y… $\bar{Y}$…

$$\textbf{SC}_{\textbf{trat}} = \frac{43.10^2 + 48^2 + 57.5^2}{(4)(2)} - \frac{(148.6)^2}{(4)(3)(2)} = 12.92$$

$$\textbf{SC}_{\textbf{ee}} = \frac{11.30^2 + 10.1^2 + ... + 12.40^2}{2} - \frac{43.10^2 + 48^2 + 57.5^2}{(4)\,(2)} = 15.25$$

$$\textbf{SC}_{\textbf{em}} = 5.6^2 + 5.72 + 5.0^2 + ... + 6.7^2 - \frac{11.30^2 + 10.1^2 + ... + 12.40^2}{2} = 3.555$$

$$\textbf{SC}_{\textbf{total}} = 5.6^2 + 5.7^2 + 5.0^2 + ... + 6.7^2 - \frac{(148.6)^2}{(4)\,(3)\,(2)} = 31.73$$

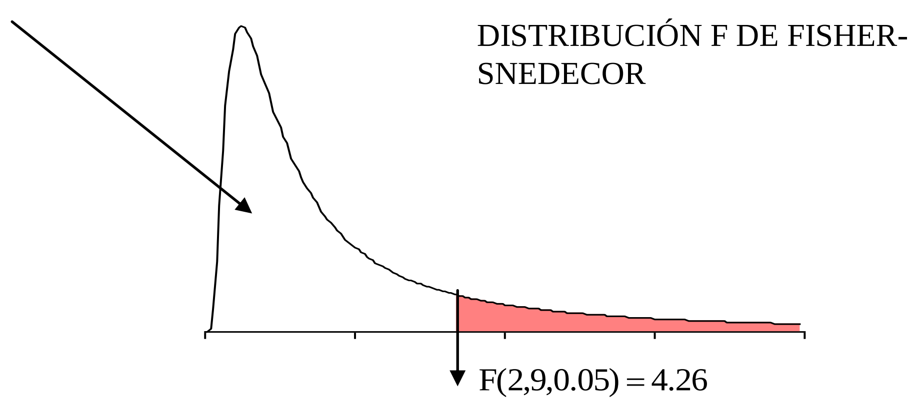
RESUMEN DEL ANÁLISIS DE VARIANZA

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados Medios	Valor de F	F crítico (5%)
Tratamientos	2	12.92	6.46		
Error Experimental	9	15.25	1.69	5.72*	2.80
Error de Muestreo	12	3.55	0.296		
Total	23	31.73			

* significativo al 0.05

Como F_1 = 5.72 es > F crítica (9,12,0.05) = 2.80, se tiene evidencia para rechazar Ho: $\sigma_e^2 = 0$, concluyendo que el muestreo fue efectivo en este experimento. Y por lo tanto se procede a calcular F_2 de la manera habitual, utilizando el CMee:

$$F_2 = \frac{CM_{trat}}{CM_{ee}} = \frac{6.4629}{1.6943} = 3.8144$$



Conclusión:

Como $F_2 < F_{t(2,9,0.05)}$: se acepta Ho, por lo tanto todos los tratamientos producen el mismo efecto.

$$CV = \frac{\sqrt{CMee}}{\bar{Y}...} \times 100 \qquad CV = \frac{\sqrt{1.71}}{6.19} \times 100 = 21.13\%$$

Nota: para obtener el valor p (*p value*) en MS Excel®, ingrese en el menú INSERTAR, busque función (***fx***) y luego seleccione la categoría **Estadísticas**, y dentro de éstas DISTR.F. (X, grados_de_libertad1, grados_de_libertad2). X es el valor al que desea evaluar la función, un número no negativo. En X debe ingresar el valor de F_2 = 3.8184, grados_de_libertad1 = 2 y grados_de_libertad2 = 9. El valor p resultante = 0.06299081, que es mayor al valor de α = 0.05; concluyendo que todos los tratamientos producen el mismo efecto.

2.7 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO

```
OPTIONS nodate nonumber;
DATA dca2;
INPUT trat $ rep resp;
LABEL resp = "variable de respuesta";
CARDS;
A    1    5.6
A    1    5.7
A    2    5
A    2    5.4
A    3    5.5
A    3    5.4
A    4    5.3
A    4    5.5
B    1    6.7
B    1    8.7
B    2    4.7
B    2    3.7
B    3    5.7
B    3    6.5
B    4    6.2
B    4    5.8
C    1    7.6
C    1    7.8
C    2    7.4
C    2    7.2
C    3    7.5
C    3    7.6
C    4    5.7
C    4    6.7
;
PROC anova;
TITLE "DCA con muestreo";
CLASS trat rep; /*con muestreo*/
MODEL resp = trat rep(trat); /*rep(trat) = error experimental*/
TEST h = trat  e = rep(trat);
RUN;
PROC anova;
TITLE "mancomunando errores"; /*sin efecto de muestreo*/
CLASS trat;
MODEL resp = trat;
run;
```

Otra alternativa, es digitando luego del conjunto de datos, esta parte del programa:

```
PROC GLM DATA= dca2;
CLASS trat rep; /*con submuestreo*/
MODEL resp = trat rep(trat)/SS3; /*rep(trat) = error
                                  experimental*/
RANDOM rep(trat)/TEST; /*se define como aleatoria la
unidad experimental REP dentro de TRAT y a través de
la opción TEST se obtienen las pruebas estadísticas
correctas*/

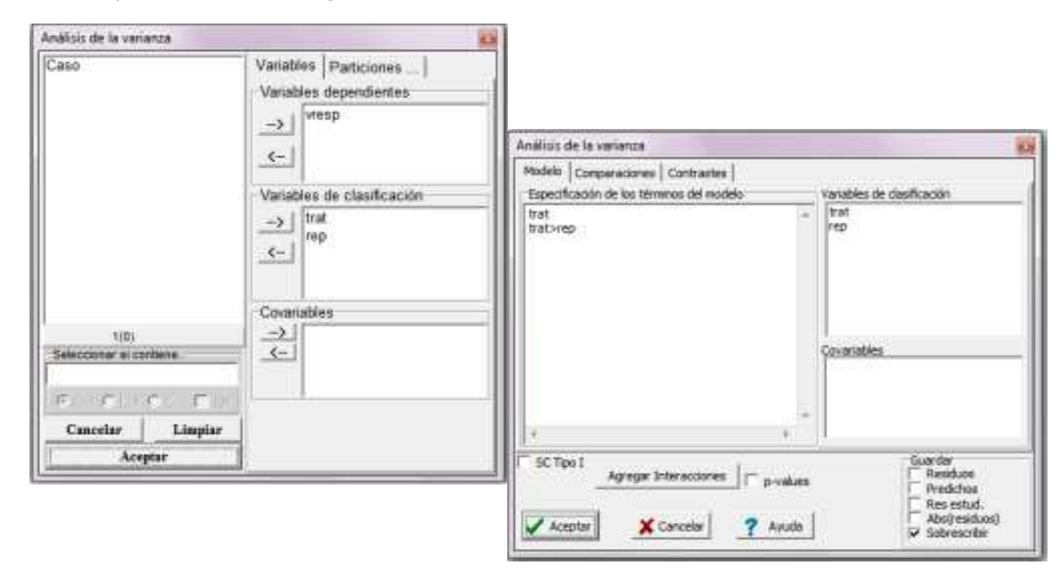
/*Análisis de Varianza para un Modelo Mixto*/
PROC MIXED DATA=dca2;
CLASS trat rep;
MODEL resp = trat; /*se colocan los efectos fijos
involucrados en el modelo*/
RANDOM rep(trat);
RUN;
```

2.8 INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO COMPLETAMENTE AL AZAR CON MUESTREO

Los datos ingresados en Infostat para el ejemplo 2.6.1 quedarían de la siguiente manera:

Caso	trat	rep	vresp
1	A	1	5,6
2	A	2	5,0
3	A	3	5,5
4	A	4	5,3
5	A	1	5,7
6	A	2	5,1
7	A	3	5,4
8	A	4	5,5
9	B	1	6,7
10	B	2	4,7
11	B	3	5,7
12	B	4	6,2
13	B	1	8,7
14	B	2	3,7
15	B	3	6,5
16	B	4	5,8
17	C	1	7,6
18	C	2	7,4
19	C	3	7,5
20	C	4	5,7
21	C	1	7,8
22	C	2	7,2
23	C	3	7,6
24	C	4	6,7

Para solicitar los resultados del análisis de varianza se sigue la ruta en el menú principal *Estadísticas / Análisis de varianza* y se definen las siguientes alternativas:



Los resultados obtenidos son los siguientes. (Compare con los resultados manuales del ANOVA del apartado 2.6.1)

F.V.	SC	gl	CM	F	p-valor
trat	12.93	2	6.46	21.82	0.0001
trat>rep	15.25	9	1.69	5.72	**0.0033***
Error	3.55	12	0.30		
Total	31.73	23			

Como el valor de F_1 = 5.72 es significativo (valor-p= 0.0033), se tiene evidencia para rechazar Ho: $\sigma_e^2 = 0$, concluyendo que el muestreo fue efectivo en este experimento. Y por lo tanto se procede a calcular F_2 de la manera habitual, utilizando el CMee.

Para ello tendrá que ejecutar de nuevo el programa, y en el cuadro de ESPECIFICACIÓN DE LOS TÉRMINOS DEL MODELO, digitar: trat\trat>rep,trat>rep.

Infostat generará el siguiente cuadro de ANOVA:

F.V.	SC	gl	CM	F	p-valor	(Error)
trat	12.93	2	6.46	3.81	0.0631^{NS}	(trat>rep)
trat>rep	15.25	9	1.69	5.72	0.0033	
Error	3.55	12	0.30			
Total	31.73	23				

Nota:

En el caso que sea necesario mancomunar los errores, en el cuadro de ESPECIFICACIÓN DE LOS TÉRMINOS DEL MODELO, solamente se debe digitar: trat. El programa calculará el valor de la estadística F para tratamientos, considerando en el denominador el cuadrado medio mancomunado.
A continuación se presenta la salida:

F.V.	SC	gl	CM	F	p-valor
trat	12.93	2	6.46	7.22	0.0041
Error	18.80	21	0.90		
Total	31.73	23			

Compare los resultados obtenidos de forma manual.

2.9 EJERCICIOS PROPUESTOS

1. Mendoza (2001) evaluó el efecto de la pulpa del café *(Coffea arabica* L.) sobre el rendimiento y eficiencia biológica de la cepa ECS-0110 de *Pleurotas ostreatus* utilizando estopa de coco *(Cocus nucifera* L.) y estróbilos de pino *(Pinus* spp.) como sustratos. Los resultados obtenidos para la variable rendimiento, expresado en gramos de hongo fresco obtenido por cada 454 gramos de sustrato seco, se presentan a continuación:

Tratamientos		Repeticiones				
		I	**II**	**III**	**IV**	**V**
T_1	Coco 100%	271.09	468.24	345.02	335.15	320.37
T_2	Coco-Pulpa 9:1	265.15	371.24	291.68	318.46	318.30
T_3	Coco-Pulpa 8:2	254.09	171.39	314.33	265.67	285.78
T_4	Coco-Pulpa 7:3	433.68	278.76	309.75	278.80	309.78
T_5	Coco-Pulpa 6:4	472.91	439.14	489.80	371.60	448.25
T_6	Coco-Pulpa 5:5	372.59	484.37	465.79	447.14	484.39
T_7	Pino 100%	147.64	210.93	164.52	147.62	189.84
T_8	Pino-Pulpa 9:1	227.30	231.85	181.83	195.48	215.92
T_9	Pino-Pulpa 8:2	197.14	234.13	189.77	246.46	246.46
T_{10}	Pino-Pulpa 7:3	349.87	296.03	376.77	242.20	376.77
T_{11}	Pino-Pulpa 6:4	355.67	385.31	355.66	281.57	385.28
T_{12}	Pino-Pulpa 5:5	527.66	428.75	346.29	303.43	362.79
T_{13}	Pulpa 100% (Testigo)	565.82	615.15	552.05	580.26	605.12

Los tratamientos son expresados en proporciones de sustrato en peso seco.

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema

2. Sosa Leonardo (1999) evaluó cuatro substancias diluyentes-dispersantes de polen para producir semilla híbrida en cuatro cultivares de Marigold *(Tagetes erecta* L.) mediante polinización artificial en condiciones de invernadero. A continuación se presentan los datos de campo de las variables: cantidad de aquenios y peso de aquenios (en gramos) medidas en Tagetes erecta P-702-1 Discovery Orange.

Tratamientos	**Cantidad de aquenios**			**Peso de Aquenios (gramos)**		
	I	**II**	**III**	**I**	**II**	**III**
Testigo	912	830	835	1.80	1.50	1.50
Leche 25%	712	671	630	1.70	1.41	1.20
Leche 50%	690	828	759	1.20	1.48	1.34
Leche 75%	656	673	769	1.50	1.68	1.70
Gelatina 25%	519	535	550	1.31	1.04	1.35
Gelatina 50%	658	635	611	1.70	1.31	1.66
Gelatina 75%	888	806	723	1.85	1.81	1.64
Harina Arroz 25%	240	195	120	1.00	0.90	0.75
Harina Arroz 50%	218	160	190	0.98	0.40	0.55
Harina Arroz 75%	120	102	135	0.32	0.25	0.70
Portulaca 25%	655	628	601	1.33	1.04	1.18
Portulaca 50%	750	1123	999	1.68	1.83	2.35
Portulaca 75%	570	595	545	1.11	1.40	1.22

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA para cada variable y concluya en términos del problema

3. Se realizó un experimento para evaluar el efecto de la adición de compuestos vitamínicos al alimento balanceado en la ganancia de peso en cerdos. Tres diferentes compuestos fueron evaluados (A, B y C) y un control (D – sin la adición de compuesto vitamínico). El aumento de peso tras una semana en una muestra aleatoria de 22 cerdos se da a continuación:

Variable de respuesta: aumento de peso (en libras) tras una semana.

A	11.1	10.9	10.8	10.6	11.4	10.7
B	11.5	11	10.8	10.2	11.2	10.9
C	10.1	10.6	11.2	9.7	10.4	9.5
D	9.2	9.8	10.1		10.4	

a) Describa el modelo estadístico-matemático.
b) Cite los supuestos del análisis de varianza, así como el nombre de las pruebas estadísticas que se utilizan para evaluarlos. ¿Qué es lo que se hace cuando no se cumplen los supuestos?
c) Plantee las hipótesis
d) Realice el análisis de varianza

4. Se realizó un experimento con el propósito de determinar si existen diferencias entre cinco variedades de menta (*Mentha spicata* var. *crispata*), denominadas: A, B, C, D y E, para lo cual se sembró cada variedad en tres macetas y se midieron los crecimientos (en cm.) en una semana, de los tallos de cuatro plantas por maceta. Los resultados se presentan a continuación:

		Número de planta			
Variedad	Maceta	1	2	3	4
	1	5.0	5.5	4.0	3.5
A	2	3.5	3.5	3.0	4.0
	3	4.5	4.0	4.0	5.0
	1	5.0	4.5	5.0	4.5
B	2	5.5	6.0	5.0	5.0
	3	5.5	4.5	6.5	5.5
	1	8.5	6.0	9.0	8.5
C	2	6.5	7.0	8.0	6.5
	3	7.0	7.0	7.0	7.0
	1	6.0	5.5	3.5	7.0
D	2	6.0	8.5	4.5	7.5
	3	6.5	6.5	8.5	7.5
	1	7.0	9.0	8.5	8.5
E	2	6.0	7.0	7.0	7.0
	3	11	7.0	9.0	8.0

Realice el ANOVA y concluya, presente en forma ordena el procedimiento.

5. Un experimento fue realizado con la finalidad de comparar cuatro líneas avícolas (tratamientos), dos especializadas en producción de carne y dos de doble propósito, los cuales se asignaron en forma aleatoria a cuatro corrales dentro de cada granja (unidad experimental). Las observaciones se hicieron en corrales individuales (muestra) y se midió la conversión alimenticia de las aves. Los datos para las líneas de producción se presentan a continuación:

Engorda								Doble propósito							
A				B				C				D			
No.de granja				No.de granja				No.de granja				No.de granja			
1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
2.5	2.3	2.1	2.0	2.5	1.8	2.0	1.8	3.5	3.5	4.0	4.0	5.5	4.0	4.0	5.0
2.3	2.0	2.5	2.0	2.5	2.0	1.7	1.9	4.0	4.0	3.0	3.5	4.5	5.0	5.0	4.0
2.2	2.0	2.4	2.5	2.0	2.0	2.0	2.0	4.0	4.3	3.5	3.5	4.5	5.5	5.0	4.5
2.4	2.5	2.0	2.0	2.0	1.9	1.9	2.0	3.5	3.5	3.0	4.0	5.5	5.5	5.5	5.0

Realice el análisis de varianza con un nivel de 5% de significancia.

6. En un ensayo de campo se incluyó cinco tratamientos de fertilización para evaluar su efecto sobre el rendimiento de cebada (*Hordeum vulgare*). De las 30 parcelas experimentales homogéneas que se disponía, se asignaron al azar seis a cada tratamiento. Al momento de la cosecha se tomaron al azar tres cuadros muestra en cada parcela, cuyos resultados (codificados) se presentan en el siguiente cuadro.

Repetición	Tratamiento de fertilizante				
	1	2	3	4	5
I	57	67	95	102	123
	46	72	90	88	101
	28	66	89	109	113
$Y_{i1.}$					
II	26	44	92	96	93
	38	68	89	89	110
	20	64	106	106	115
$Y_{i2.}$					
III	39	57	91	102	112
	39	61	82	93	104
	43	61	98	98	112
$Y_{i3.}$					
IV	23	74	105	103	120
	36	47	85	90	101
	18	69	85	105	111
$Y_{i4.}$					
V	48	61	78	99	113
	35	60	89	87	109
	48	75	95	113	111
$Y_{i5.}$					
VI	50	68	85	117	124
	37	65	74	93	102
	19	61	80	107	118
$Y_{i6.}$					
$Y_{i..}$					
$Y_{i1.}$					

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema.
d) Comente sobre la efectividad de usar submuestreo.

7. En el curso de Industrialización de la Madera fue realizado un ensayo de tres productos químicos tendientes a retardar la expansión del fuego cuando es usado en el tratamiento de páneles para piso de madera. El investigador obtuvo 12 páneles y aplicó cada uno de los productos a cuatro de ellos. Para tener mayor precisión, cada pánel fue cortado en dos piezas y luego midió el tiempo requerido (minutos) por cada uno de ellos para ser consumido por el fuego. Los resultados se resumen en la siguiente tabla:

Pánel (Repetición)	Muestra	Producto Químico A	Producto Químico B	Producto Químico C
1	1	10.3	4.4	3.1
	2	9.8	4.7	3.3
		20.1	**9.1**	**6.4**
2	1	5.8	2.7	6.5
	2	5.4	1.6	5.4
		11.2	**4.3**	**11.9**
3	1	8.7	4.6	5.1
	2	10.0	4.0	7.5
		18.7	**8.6**	**12.6**
4	1	8.9	5.6	5.6
	2	9.4	3.4	4.2
		18.3	**9.0**	**9.8**
	Yi..	68.3	31.0	40.7

e) Plantee las hipótesis a evaluar.
f) Describa el modelo estadístico matemático
g) Realice el ANOVA y concluya en términos del problema.
h) Comente sobre la efectividad de usar submuestreo.

8. Banzatto y Kronka (2011) citan los resultados obtenidos por Cardoso Filho (1974) [1/], referentes a la evaluación de 5 cultivares de sorgo (*Sorghum* spp.), la variable de respuesta medida fue la producción de materia seca, expresada en tm ha^{-1}. Los datos se presentan a continuación:

Cultivares	R_1	R_2	R_3	R_4	R_5	R_6
NK 300 (híbrido)	10.3	11.6	11.7	11.4	11.2	11.2
Sordan 67 (híbrido)	9.8	10.0	10.2	11.9	10.4	10.5
Pioneer 988 (híbrido)	9.9	9.6	10.0	10.4	---	---
Pioneer 93 (híbrido)	21.2	20.6	22.3	19.9	21.0	---
SART (Variedad)	20.2	20.6	22.1	20.8	20.9	20.9

[1/] Cardoso Filho, AR. 1974. Competição de sorgos forrageiros (*Sorghum bicolor* L. Moench) na região de Jaboticabal. Rendimento de massa verde, matéria seca e composição bromatológica da silagem. Trabalho de conclusão de Curso (Graduação em Agronomia) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal. 40 p.

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema.

9. Un experimento fue realizado para probar el efecto de cinco fuentes de energía utilizadas en dietas para engorda de toretes (T_1. Testigo, T_2. Melaza, T_3. Cebo, T_4. Maíz, T_5. Sorgo) en las cuales se midió la ganancia de peso (GP) durante el período de engorda. Se consideraron 5 repeticiones por tratamientos (25 animales) y se planteó la hipótesis de igualdad de medias de tratamientos.

	Tratamientos				
Repetición	T_1	T_2	T_3	T_4	T_5
1	980	1200	1300	1400	1350
2	1050	1230	1180	1350	1420
3	1100	1150	1200	1380	1550
4	1000	1390	1170	1420	1600
5	1120	1250	1050	1500	1490

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema.

10. En un experimento reportado por Banzatto y Kronka (2011) realizado bajo un diseño completamente al azar, fueron evaluados 5 cultivares de yuca (*Manihot esculenta* L.):

A. IAC 5 **B.** IAC 7 **C.** IAC 11 **D.** Iracema y **E.** Mantiqueira.

IAC = Instituto Agronómico de Campinas (www.iac.sp.gov.br)

La asignación de los tratamientos a las parcelas en el campo, junto con las producciones en tm ha^{-1} se presentan en la figura siguiente:

(A_3) 20.3	(E_1) 47.8	(C_3) 25.8	(B_5) 28.7	(B_1) 20.9
(B_4) 28.3	(D_2) 43.2	(A_5) 29.3	(A_1) 38.9	(D_3) 41.7
(E_2) 47.8	(A_2) 25.4	(E_4) 50.5	(D_1) 38.7	(C_1) 28.1
(C_2) 27.0	(D_5) 40.3	(B_3) 32.3	(C_4) 26.9	(B_2) 26.2
(E_5) 56.4	(A_4) 25.7	(C_5) 22.3	(E_3) 44.7	(D_4) 39.0

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema.

11. Planifique un experimento, utilizando un diseño completamente al azar. Describa los tratamientos, número de repeticiones, justifique el uso de este diseño. Haga el croquis de campo y muestre la aleatorización.

12. Revise tres trabajos de investigación agropecuaria, en los cuales haya sido utilizado el diseño completamente al azar con muestreo. Presente el cuadro de resumen de los datos y analícelos siguiendo el procedimiento explicado en el texto.

13. *Montana Gourmet Garlic* es una empresa que se dedica al cultivo de ajo (*Allium sativum*) utilizando métodos orgánicos. Esta empresa se especializa en las variedades tipo *hardneck*. Los propietarios diseñaron un experimento para evaluar si el crecimiento del ajo es afectado por el tipo de fertilizante utilizado. En el experimento fue utilizada una variedad de ajo Rocambole llamada Spanish Roja y evaluados tres abonos orgánicos y un fertilizante químico (como control). Un acre de tierra cultivable fue reservado para el experimento, y dividido en 32 camellones, y asignados de forma aleatória los tipos de fertilizante. Al momento de la cosecha, fueron calculados los pesos promedios de bulbos (en onzas) de ajo en cada camellón y el número promedio de dientes de ajo en cada bulbo. Los resultados se presentan en la tabla siguiente:

Fertilizante	Peso promedio	No. dientes/bulbo	Fertilizante	Peso promedio	No. dientes/bulbo
1	0.24402	12.5793	3	0.25792	12.255
1	0.20891	11.5416	3	0.2015	10.6891
1	0.23277	11.1439	3	0.2335	11.213
1	0.20161	13.1160	3	0.21481	11.2933
1	0.25285	12.0435	3	0.22271	11.4917
1	0.25362	11.4838	3	0.23229	12.5118
1	0.24118	12.5481	3	0.18914	12.1337
1	0.2053	12.0564	3	0.21213	11.8415
1	0.1884	13.0964	3	0.25785	10.1495
2	0.21588	12.0982	3	0.27828	10.7842
2	0.24114	9.9072	3	0.2279	11.7475
2	0.1956	10.0428	4	0.20901	11.5062
2	0.15851	12.4579	4	0.24754	12.9199
2	0.22152	11.6716	4	0.15173	14.0173
2	0.23194	10.9575	4	0.17713	12.6372
2	0.18979	12.0743			
2	0.21414	12.7071			

Fuente: Huber, M. 2013. **SAS® Enterprise Guide®: Regression and ANOVA for Professors Course Notes**. Cary, NC: SAS Institute Inc. 123 p. Curso recibido por el Dr. Sc. Ezequiel López en el CIAGRI, ESALQ, USP, 9-11 de noviembre de 2013.

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA para las dos variables y concluya en términos del problema.
d) Consideranto que el tratamiento 4 es el control, revise en el Capítulo 3, como realizar la prueba de Dunnett.

14. Zamudio y Alvarado (1996) citan un experimento en que los tratamientos consistieron en evaluar combinaciones de temperatura y tiempo en el proceso de prensado de tableros de madera; las unidades experimentales eran los tableros de aglomerado que se remojaron en agua durante dos horas para estudiar su respuesta de hinchamiento en relación a su dimensión original. La medición de la variable respuesta (porcentaje de hinchamiento), se realizó en submuestras denominadas para este caso, probetas, que son subdivisiones del tablero. El diseño experimental utilizado fue completamente al azar, con 2 repeticiones. Los tratamientos se describen a continuación:

Tratamiento	Temperatura en °C	Tiempo de prensado
A	150	6.0 minutos
B	150	8.5 minutos
C	220	6.0 minutos
D	220	3.5 minutos

Los porcentajes de hinchamiento de los tableros se presentan en el siguiente cuadro:

Tratamientos							
A		**B**		**C**		**D**	
Rep1	**Rep2**	**Rep1**	**Rep2**	**Rep1**	**Rep2**	**Rep1**	**Rep2**
42.98	46.25	32.59	31.84	21.87	21.08	41.49	40.62
40.26	43.30	30.66	29.73	19.64	20.18	33.48	33.63
38.49	38.84	33.33	29.41	20.53	17.11	32.16	27.03
36.40	35.87	28.13	28.51	20.53	21.62	32.31	32.43
40.43	37.77	31.25	31.08	21.24	19.00	32.47	27.17
41.88	42.04	32.59	31.83	23.55	20.72	40.77	38.39
42.29	40.62	34.80	30.94	22.22	19.28	33.19	36.61
38.59	37.38	28.76	28.05	20.09	18.83	29.65	31.39
37.55	34.68	29.77	27.27	21.97	18.30	29.20	28.38

Fuente: Zamudio, F.; Alvarado, A. 1996. **Análisis de Diseños Experimentales con igual número de submuestras**. Universidad Autónoma Chapingo, División de Ciencias Forestales. 58 p. Disponible en: http://www.geocities.ws/a_alvaseg/submuestreov1.pdf; consultado el 20/12/2013.

a) Plantee las hipótesis a evaluar.
b) Describa el modelo estadístico matemático
c) Realice el ANOVA y concluya en términos del problema.
d) Resuelva el ejercicio utilizando Infostat.

CAPÍTULO 3
PRUEBAS DE COMPARACIÓN MÚLTIPLE DE MEDIAS

3.1 INTRODUCCIÓN

Para poder tener una mayor comprensión sobre este tema, inicialmente se responderán algunas interrogantes:

a) ¿Para qué se utiliza un análisis posterior al análisis de varianza?

Se requiere del uso de algún método de análisis posterior al ANOVA para contrastar diferentes subhipótesis de interés, después que se verifica que el valor de la estadística F para alguna de las hipótesis en la tabla de ANOVA es significativa. Cada una de las hipótesis que se rechaza en la tabla del ANOVA, comprobada por el valor crítico respectivo de F, le corresponde una o varias subhipótesis que se deben contrastar por un método apropiado de análisis posterior.

b) ¿Qué métodos de análisis posterior existen?

b.1) Pruebas de comparación múltiple de medias, de acuerdo con los criterios de:

- Tukey (1953)
- Duncan (1955)
- SNK (Student-Newman-Keuls). Diseñada por Newman (1939) y estudiada por Keuls (1952)
- Bonferroni (el uso moderno de esta prueba es atribuído a Dun, O.J., 1961)
- Scheffé (1953)
- Dunnett (1955, 1964)
- Scott Knott (1974), entre otras.

b.2) Contrastes lineales ortogonales y no ortogonales

b.3) Polinomios ortogonales.

Nota: El nivel de significancia que ha sido utilizado para determinar la significancia de un valor de F en la tabla de ANOVA, es el que debe ser utilizado para el análisis posterior.

Estos métodos se aplican **"regularmente"** cuando la hipótesis de igualdad de las medias de los tratamientos en el análisis de varianza ha sido rechazada. A continuación se presenta la descripción de algunas pruebas de comparación múltiple de medias.

3.2 COMPARACIÓN MÚLTIPLE DE MEDIAS, SEGÚN EL CRITERIO PROPUESTO POR TUKEY

Llamado de rango estudiantizado, construido por Tukey en 1953, y conocido como la prueba de la diferencia significativa honesta HSD (Honestly significant difference), este método sirve para comparar las medias de los tratamientos, dos a dos, o sea, para evaluar las hipótesis:

Ho: $\mu_i = \mu_j$ (media del tratamiento i es igual a la media del tratamiento j, con $i \neq j$)
Ha: $\mu_i \neq \mu_j$ (media del tratamiento i es diferente a la media del tratamiento j, con $i \neq j$)

Caso I: DCA Balanceado

Ejemplo: El análisis de varianza que a continuación se presenta, corresponde a un experimento realizado en arroz *(Oryza sativa* L.), en el que se evaluó nueve insecticidas para el control de larvas de una determinada plaga. La variable de respuesta medida fue el número de larvas vivas, a la cual se le aplicó la transformación raíz cuadrada. El diseño experimental utilizado fue completamente al azar, con 4 repeticiones. El cuadro resumen del ANOVA se presenta a continuación:

Fuentes de variación	**Grados de libertad**	**Suma de cuadrados**	**Cuadrados Medios**	**F**	**Valor crítico de F**
Insecticidas	8	46.04	5.76	4.26*	2.31
Error Experimental	27	36.43	1.35		
Total	35	82.48			

* significativo al 5%

Debido a que se detectaron diferencias significativas en el efecto de los insecticidas, se aplicará la prueba de comparación múltiple de medias de acuerdo con el criterio de Tukey. El procedimiento se detalla a continuación:

1. Se deben obtener las medias de los tratamientos

Insecticida	Medias (*)
1	1.87
2	2.02
3	2.55
4	2.42
5	3.91
6	4.12
7	0.25
8	1.39
9	2.82

(*) Las medias fueron obtenidas a partir de los datos transformados.

2. Se construye una matriz de diferencias entre todos los posibles pares de medias

Insecticida	medias	6	5	9	3	4	2	1	8	7
		4.12	3.91	2.82	2.55	2.42	2.02	1.87	1.39	0.25
7	0.25	3.87	3.66	2.57	2.3	2.17	1.77	1.62	1.14	–
8	1.39	2.73	2.52	1.43	1.16	1.03	0.63	0.48	–	
1	1.87	2.25	2.04	0.95	0.68	0.55	0.15	–		
2	2.02	2.1	1.89	0.8	0.53	0.4	–			
4	2.42	1.7	1.49	0.4	0.13	–				
3	2.55	1.57	1.36	0.27	–					
9	2.82	1.3	1.09	–						
5	3.91	0.21	–							
6	4.12	–								

Cada una de las diferencias ($d_{ii'}$) fueron obtenidas con la siguiente ecuación:

$$d_{ii'} = \left| \overline{Y}_{i.} - \overline{Y}_{i.'} \right| \text{ , siendo que } i \neq i'$$

3. Se calcula W, la diferencia mínima significativa a un cierto nivel de significancia (α), dada por la siguiente expresión:

$$W = q_{(t, glee, \alpha)} \times \sqrt{\frac{CMee}{r}}$$

siendo:

q = amplitud total estudentizada. Valor encontrado en tablas y que está en función de:
α. = (nivel de significancia)
t = (número de tratamientos), y
glee = (grados de libertad del error experimental)
CMee = cuadrado medio del error experimental
r = número de repeticiones de las medias de los tratamientos a ser comparadas.

Para nuestro ejemplo tenemos que al consultar la Tabla 2 del Apéndice, se obtiene:

$$q_{(9,27,0.05)} = 4.774$$

Como en la tabla no se encuentra el valor exacto de q, se efectuó una interpolación:

glee	q
20	4.90
27	X
30	4.72

10 (30 – 20); 27–20 = 7

4.90 – 4.72 = 0.18

10 – 0.18
7 – X

X = (7 × 0.18) / 10 = 0.126
q = 4.90 – 0.126 = 4.774

$$W = 4.774 \times \sqrt{\frac{1.349}{4}} = 2.7724$$

4. Volvemos a la matriz de diferencias (Paso 2) y observamos columna por columna, si $d_{ii'} \geq W$, significa que existen diferencias significativas entre los efectos de los pares de tratamientos, y colocamos un asterisco para resaltar esas diferencias.

Insecticida	medias	6 4.12	5 3.91	9 2.82	3 2.55	4 2.42	2 2.02	1 1.87	8 1.39	7 0.25
7	0.25	3.87*	3.66*	2.57	2.3	2.17	1.77	1.62	1.14	–
8	1.39	2.73	2.52	1.43	1.16	1.03	0.63	0.48	–	
1	1.87	2.25	2.04	0.95	0.68	0.55	0.15	–		
2	2.02	2.1	1.89	0.8	0.53	0.4	–			
4	2.42	1.7	1.49	0.4	0.13	–				
3	2.55	1.57	1.36	0.27	–					
9	2.82	1.3	1.09	–						
5	3.91	0.21	–					**W=2.774**		
6	4.12	–								

5. Presentación de los resultados

Insecticida	Medias	Grupo Tukey
6	4.12 (20)	a
5	3.91 (16)	a
9	2.82 (8.50)	a b
3	2.55 (7.25)	a b
4	2.42 (6)	a b
2	2.02 (4.5)	a b
1	1.87 (5.5)	a b
8	1.39 (3.25)	a b
7	0.25 (0.25)	b

Entre ( ) aparecen los promedios de los datos originales (sin transformar).

Conclusión: la aplicación del insecticida 7 presenta los mejores resultados en cuanto al control de larvas.

Caso II: **DCA No balanceado**

Se tomarán los datos del ejemplo desarrollado en la clase.

1. Matriz de diferencias

Suplemento	Media (r_i)	Araruta 34.40 (6)	Mandioca 23.64 (5)	Batata doce 22.84 (6)	Sin Supl. 22.64 (6)
Sin Suplemento	22.64 (6)	11.76 *	1	0.2	–
Batata Doce	22.84 (6)	11.56 *	0.8	–	
Mandioca	23.64 (5)	10.76 *	–		
Araruta	34.40 (6)	–			

2. Comparador de Tukey

a) Entre medias con el mismo número de repeticiones, esto es $r_i = r_{i'}$, para $i \neq i'$

$$W = q_{(4,19,0.05)} \times \sqrt{\frac{CMee}{r}}$$

$$W = 3.98 \times \sqrt{\frac{3.53}{6}} = 3.05$$

b) Entre medias con número diferente de repeticiones, esto es: $r_i = 5$ y $r_{i'} = 6$, para $i \neq i'$

$$W' = q_{(4,19,0.05)} \times \sqrt{\frac{1}{2} \times CMee \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)}$$

$$W' = 3.98 \times \sqrt{\frac{1}{2} \times 3.53 \left(\frac{1}{5} + \frac{1}{6} \right)} = 3.20$$

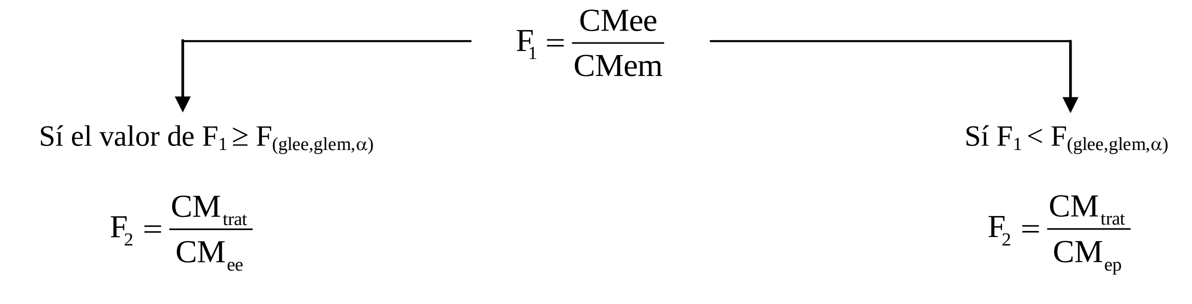
c) Presentación de los resultados

Suplemento	**Producción media diaria de leche (kg)**	**Grupo Tukey**
Araruta	34.40	a
Mandioca	23.64	b
Batata doce	22.84	b
Sin Supl.	22.64	b

Se concluye que el suplemento alimenticio que proporcionó la mayor producción diaria (kg) de leche fue el suplemento preparado con la planta Araruta.

Caso III: DCA con submuestreo

Recuerde que primero se realiza una prueba de F con el error experimental y el error de muestreo:



Valor crítico de $F_{(glt,glee,\alpha)}$

Valor crítico de $F_{(glt,glep,\alpha)}$

$$W = q_{(t,glee,\alpha)} \times \sqrt{\frac{CMee}{r}}$$

$$W = q_{(t,glep,\alpha)} \times \sqrt{\frac{CMep}{rm}}$$

Nogueira (2007) indica algunas consideraciones referentes a este método:

a) El método Tukey fue basado en la distribución de la diferencia entre la menor y la mayor estadística de orden (*range*) de una muestra;

b) Este método es válido en la totalidad de los contrastes de medias, dos a dos;

c) El método de Tukey es exacto, cuando los tratamientos están balanceados;

d) El método de Tukey es exacto para evaluar la mayor diferencia entre dos medias, en los demás casos es conservador.

3.3 COMPARACIÓN MÚLTIPLE DE MEDIAS, SEGÚN EL CRITERIO DE DUNCAN

Un procedimiento usado ampliamente para comparar todas las parejas de medias es el de la prueba de intervalos múltiples desarrollada por Duncan (1955). La aplicación de esta prueba es más laboriosa que la prueba de Tukey, pero se llega a resultados más detallados y se discrimina con mayor facilidad entre los tratamientos, o sea que, la prueba de Duncan indica resultados significativos en casos en que la prueba de Tukey no permite obtener significancia estadística. Tal como la prueba de Tukey, la de Duncan exige, para ser exacto, que todos los tratamientos tengan el mismo número de repeticiones.

Para el uso de esta prueba se necesitan tablas especiales (Tabla 3). A continuación se presenta un ejemplo ilustrativo.

Orellana Najarro (2006) evaluó la selectividad de los herbicidas Acetoclor y Alaclor en seis cultivos hortícolas en el municipio de Monjas, Jalapa. Una de las variables utilizadas fue altura de planta (cm) 30 días después del transplante. El diseño experimental utilizado fue bloques al azar, con 3 repeticiones; el cuadrado medio del error = 18.8827 y los grados de libertad = 10. Los promedios de los tratamientos en orden descendente son:

No. tratamiento	**Tratamientos**	**Altura promedio**
1	Sin herbicida (testigo limpio)	26.40
2	Alaclor aplicado 2 días antes del transplante	23.13
3	Alaclor aplicado 2 días después del transplante	22.07
4	Sin herbicida (testigo enmalezado)	21.80
5	Acetoclor aplicado 2 días antes del transplante	19.00
6	Acetoclor aplicado 2 días después del transplante	13.13

El error estándar de cada promedio es: $S_{\bar{x}} = \sqrt{\frac{\text{CMee}}{\text{r}}} = \sqrt{\frac{18.8827}{3}} = 2.5088$. Usando la Tabla 3, para 10 grados de libertad y α = 0.05, inicialmente se calcula una amplitud total mínima significativa (*shortest significative range*, en inglés) para el contraste de pares de medias, dependiendo de la distancia entre cada par. Con estos datos se calculan los t–1 comparadores, usando la ecuación:

$$D_p = d_{(p,glee,\alpha)} \times \sqrt{\frac{CMee}{r}}, \text{ p = 2, 3, . . . , t (tratamientos)}$$

Siendo:

d	=	amplitud total mínima significativa. Valor encontrado en tablas y que depende de:
α.	=	(nivel de significancia)
p	=	distancia entre dos medias comparadas, y
glee	=	(grados de libertad del error experimental)
CMee	=	cuadrado medio del error experimental
r	=	número de repeticiones de las medias de los tratamientos a ser comparadas.

Las diferencias mínimas significativas para el nivel de protección α =0.05 son las siguientes:

$D_2 = 7.90$ $\quad d_{0.05}\ (2,10) = 3.15$
$D_3 = 8.28$ $\quad d_{0.05}\ (3,10) = 3.30$
$D_4 = 8.45$ $\quad d_{0.05}\ (4,10) = 3.37$
$D_5 = 8.61$ $\quad d_{0.05}\ (5,10) = 3.43$
$D_6 = 8.68$ $\quad d_{0.05}\ (6,10) = 3.46$

En la tabla siguiente se presentan las diferencias entre las medias de las variedades confrontadas con el D respectivo.

Contrastes entre medias de tratamientos	$d_{ii'} = \lvert \overline{Y}_{i.} - \overline{Y}_{i.'} \rvert$	**Distancia entre medias**	**Comparador**	
1 – 2	3.27	2	7.90	n.s.
1 – 3	4.33	3	8.28	n.s.
1 – 4	4.60	4	8.45	n.s.
1 – 5	7.40	5	8.61	n.s.
1 – 6	13.27	6	8.68	*
2 – 3	1.06	2	7.90	n.s.
2 – 4	1.33	3	8.28	n.s.
2 – 5	4.13	4	8.45	n.s.
2 – 6	10.00	5	8.61	*
3 – 4	0.27	2	7.90	n.s.
3 – 5	3.07	3	8.28	n.s.
3 – 6	8.94	4	8.45	*
4 – 5	2.80	2	7.90	n.s.
4 – 6	8.67	3	8.28	*
5 – 6	5.87	2	7.90	n.s.

La presentación final queda de la siguiente forma:

	Tratamientos	**Altura promedio**	**Grupo Duncan**
1	Sin herbicida (testigo limpio)	26.40	a
2	Alaclor aplicado 2 días antes del transplante	23.13	a
3	Alaclor aplicado 2 días después del transplante	22.07	a
4	Sin herbicida (testigo enmalezado)	21.80	a
5	Acetoclor aplicado 2 días antes del transplante	19.00	a b
6	Acetoclor aplicado 2 días después del transplante	13.13	b

3.4 MÉTODO DE DUNNETT

En varias ocasiones se ejecutan experimentos, en los que el objetivo principal es comparar determinados tratamientos con un control o testigo, siendo las comparaciones entre los demás tratamientos de interés secundario. Así, este método es recomendado cuando se desea evaluar un contraste de tipo: $Y = \mu_{i.} - \mu_a$,

donde: μ_a = se refiere a la media poblacional del tratamiento testigo o control, y
$\mu_{i.}$ = se refiere a la media poblacional del i-ésimo tratamiento o nivel del factor.

Las hipótesis a ser evaluadas son:

$$H_0 : \mu_i - \mu_a = 0, \quad o \quad H_0 : \mu_i = \mu_a,$$

contra

$$H_a : \mu_i - \mu_a \neq 0, \quad o \quad H_a : \mu_i \neq \mu_a,$$

para i = 1, . . . , a–1. El procedimiento de Dunnett (1964) es una modificación de la prueba de t. Para cada hipótesis se calculan las diferencias que se observan en las medias muestrales:

$$\left| \bar{y}_{i.} - \bar{y}_a \right|, \; i = 1,2 \; . \; . \; . \; , a-1.$$

La hipótesis nula $H_0 : \mu_i - \mu_a = 0$ es rechazada con un nivel de error tipo I según α si:

$$\left| \bar{y}_{i.} - \bar{y}_a \right| > d_\alpha(a-1, gle)\sqrt{\frac{2\ CMee}{r}}$$, y en el caso de ser desbalancedado:

$$\left| \bar{y}_{i.} - \bar{y}_a \right| > d_\alpha(a-1, gle)\sqrt{CMee\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right)},$$

en donde la constante $d_\alpha(a-1, gle)$ se encuentra en la Tabla 4 del Apéndice (son posibles tanto pruebas unilaterales como bilaterales). Hay que notar que α representa el nivel de significancia conjunto asociado a las a–1 pruebas.

Ejemplo:

Se evaluaron 4 variedades de caña de azúcar: CP-722086 (utilizada como testigo), 1, 2 y 3; las variables de respuesta medidas fueron: toneladas de caña por hectárea (TCH) y libras de azúcar por tonelada de caña (LATC). En el ensayo se utilizó un diseño de bloques completos al azar, con 5 repeticiones. Para la variable TCH se presenta a continuación el análisis de varianza:

FV	**GL**	**SC**	**CM**	**Fo**	**F crítica**
Variedades	3	1637.00	545.67	3.71	3.49
Bloques	4	103.30			
Residuos	12	1763.50	146.96		
Total	19	3503.80			

CV = 10.20%

En este ejemplo, a = 4, a–1 = 3, glee = 12, r = 5, y con un nivel de 5% de significancia se encuentra en la Tabla 4 del Apéndice que $d_{0.05}(3,12) = 2.68$. Por lo tanto la diferencia crítica es:

$$d_{0.05}(3,12)\sqrt{\frac{2\times 146.96}{5}} = 2.68\times 7.67 = 20.55$$

En consecuencia, una variedad debe considerarse significativamente diferente del control sí la diferencia es mayor que 20.55. Las diferencias observadas entre las medias son:

Variedad 1 vs. CP-722086: $\bar{y}_1 - \bar{y}_a$ = 132.00 – 109.8 = 22.20 *

Variedad 2 vs. CP-722086: $\bar{y}_2 - \bar{y}_a$ = 111.20 – 109.8 = 1.40

Variedad 3 vs. CP-722086: $\bar{y}_3 - \bar{y}_a$ = 122.60 – 109.8 = 12.80

Sólo la diferencia $\bar{y}_1 - \bar{y}_a$ indica una diferencia significativa al ser comparada con el testigo (control); por lo tanto se concluye que $\mu_1 \neq \mu_a$

Ejercicio:

Relacionado con el ejercicio anterior, realice la prueba de Dunnett, usando un 5% de significancia, para la variable LATC. Los resultados del ANOVA se presentan a continuación:

FV	**GL**	**SC**	**CM**	**Fo**	**Fcrítica**
Variedades	3	2978.80	992.93	3.85	3.49
Bloques	4	2373.7			
Resíduos	12	3094.70	257.89		
Total	19	8447.2			
				CV	5.21 %

Las medias de las variedades son:

Variedades	Promedio
CP-722086	327.6
1	303.4
2	307.6
3	294.2

3.5 EJERCICIOS PROPUESTOS

En los casos en que se presenten diferencias significativas, en los ejercicios propuestos en 2.7, realice las pruebas de comparación múltiple de medias que considere adecuadas.

3.6 PROGRAMA EN SAS PARA EL ANÁLISIS POSTANOVA DE UN EXPERIMENTO

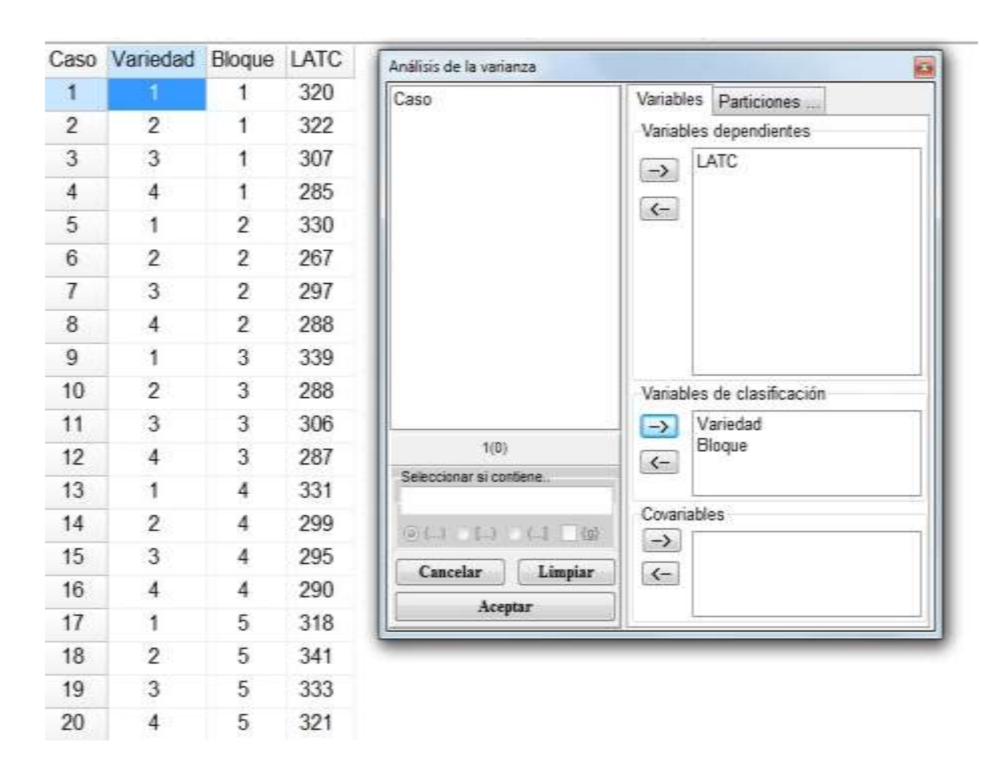
```
OPTIONS nodate nonumber;
DATA medias;
INPUT var  rep latc;
LABEL latc = "libras de azúcar por tonelada de caña”;
CARDS;
1    1    320
2    1    322
3    1    307
4    1    285
1    2    330
2    2    267
3    2    297
4    2    288
1    3    339
2    3    288
3    3    306
4    3    287
1    4    331
2    4    299
3    4    295
4    4    290
1    5    318
2    5    341
3    5    333
4    5    321
;
PROC anova;
TITLE "Análisis de varianza –prueba de Tukey";
CLASS var rep;
MODEL latc = var rep/*diseño bloques completos al azar*/
MEANS var/TUKEY;
MEANS var/TUKEY ALPHA=0.01;
RUN;

PROC anova;
TITLE "Análisis de varianza –prueba de Duncan";
CLASS var rep;
MODEL latc = var rep/*diseño bloques completos al azar*/
MEANS var/DUNCAN;
RUN;

PROC anova;
TITLE "Análisis de varianza –prueba de Dunnett bilateral";
CLASS var rep;
MODEL latc = var rep;/*diseño bloques completos al azar*/
MEANS var/DUNNETT (“1”); /*1 es la variedad testigo*/
RUN;
```

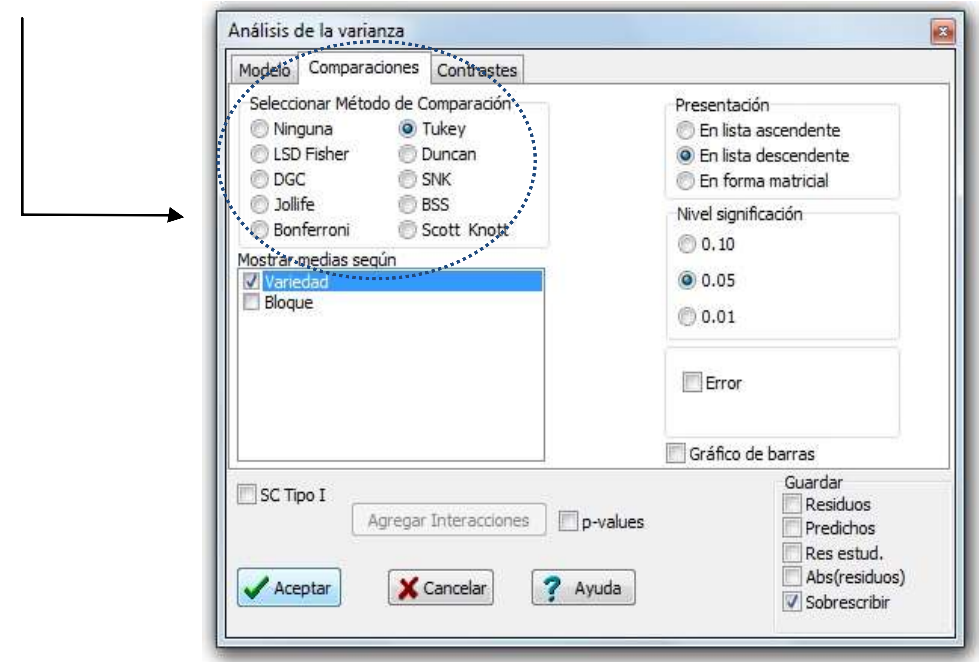
3.7 INGRESO DE DATOS EN INFOSTAT PARA EL ANÁLISIS DE UN EXPERIMENTO Y APLICACIÓN DE PRUEBAS DE COMPARACIÓN MÚLTIPLE DE MEDIAS

Los datos ingresados en Infostat para el ejemplo sobre variedadades de caña quedarían de la siguiente manera:



Caso	Variedad	Bloque	LATC
1	1	1	320
2	2	1	322
3	3	1	307
4	4	1	285
5	1	2	330
6	2	2	267
7	3	2	297
8	4	2	288
9	1	3	339
10	2	3	288
11	3	3	306
12	4	3	287
13	1	4	331
14	2	4	299
15	3	4	295
16	4	4	290
17	1	5	318
18	2	5	341
19	3	5	333
20	4	5	321

Las pruebas de comparación múltiple de medias disponibles en Infostat se presentan en el cuadro siguiente:



CAPÍTULO 4
SUPUESTOS FUNDAMENTALES DEL ANÁLISIS DE VARIANZA, VERIFICACIÓN Y TRANSFORMACIÓN DE DATOS

4.1 INTRODUCCIÓN

Para el análisis e interpretación de un conjunto de datos provenientes de un experimento, se hace uso del análisis de varianza o del análisis de regresión, considerando un modelo matemático-estadístico (Gauss-Markov), lo cual presupone un modelo lineal y la aceptación de algunas suposiciones básicas, las cuales son:

1. Los diversos efectos son aditivos. Esa condición es impuesta por el modelo adoptado. Esta suposición puede ser verificada por la prueba de Aditividad de Tukey (*Tukey test for additivity,* 1949).

2. Los errores o desvíos (e_{ij}) son independientes, esto es, la probabilidad de que el error de una observación cualquiera tenga un determinado valor, no debe depender de los valores de los otros errores. De donde resulta, que los errores no son correlacionados. El cumplimiento de esta suposición se garantiza, hasta cierto punto, por la aleatorización de los tratamientos en las unidades experimentales y mediante una buena técnica experimental (uso de borduras, evitar contagio entre unidades experimentales, etc). Puede ser utilizada la prueba de las rachas o corridas (*run test*) o la prueba de Durbin-Watson para verificar la existencia de correlación serial entre los errores. Una forma de poder tomar en cuenta la correlación espacial de parcelas, esto es, tendencia de observaciones que están en parcelas cercanas a ser más parecidas que las están más lejos, es utilizando modelos geoestadísticos que incorporen correlación espacial a nivel de los términos de error. Para más información sobre este tema puede consultar el texto: **"Aplicaciones de Modelos Mixtos en Agricultura y Forestería"** de Balzarini, Macchiavelli y Casanoves (2010).

3. Los errores (e_{ij}) tienen la misma variancia σ^2 (a esto se le conoce como *homocedasticidad* u *homogeneidad de varianzas*). Cuando las varianzas no son homogéneas, se dice que existe heterocedasticidad. La hetegoneidad de varianzas, según Banzatto y Kronka (2011) puede ser de dos tipos:

   a) Heterogeneidad irregular: ocurre cuando ciertos tratamientos presentan mayor variabilidad que otros, como en los experimentos con insecticidas, en los cuales es considerado un grupo de parcelas no tratadas (testigo). De un modo general, verificamos que los números de insectos vivos en las parcelas tratadas son menores y más homogéneos que los del testigo, que presentan mayor variabilidad.

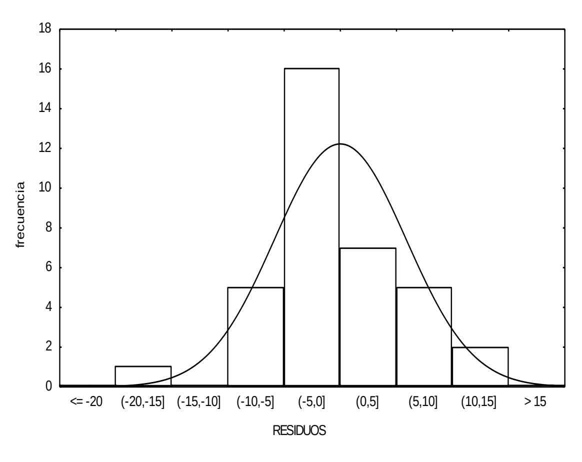
   b) Heterogeneidad regular: ocurre debido a la falta de normalidad de los datos experimentales, existiendo, frecuentemente, cierta relación entre la media y la varianza de los diversos tratamientos evaluados. Si la distribución de los datos es conocida, la relación entre media y varianza de los tratamientos también lo será, y los datos podrán ser transformados de forma que pasen a tener una distribución aproximadamente normal y las medias y varianzas se tornen independientes, permitiendo estructurar el análisis de varianza.
   Para evaluar la heterogeneidad (o la homogeneidadad) de varianzas, se utilizan las pruebas que se describen a continuación:

- Hartley (o de F máximo): es sencilla, pero se puede aplicar únicamente cuando los tamaños de muestra (repeticiones) son iguales y si los errores están normalmente distribuidos.

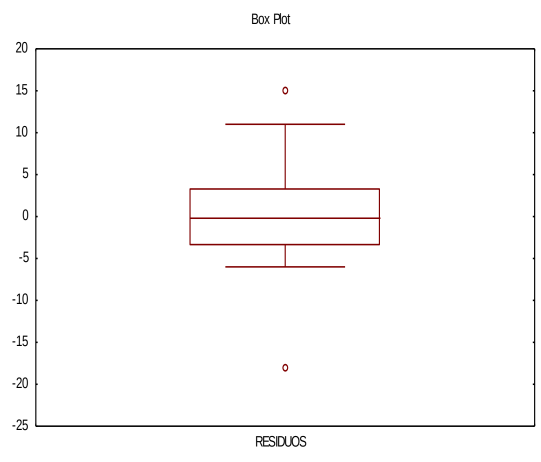
- Bartlett: esta prueba puede utilizarse cuando los tamaños de muestra (repeticiones) son iguales o diferentes. Presenta también el inconveniente de ser sensible a la falta de normalidad de los datos.

- Levene modificado: Esta prueba es útil aún cuando no se cumple con el supuesto de normalidad y además, no se requiere que los tamaños de muestra sean los mismos para todos los tratamientos.

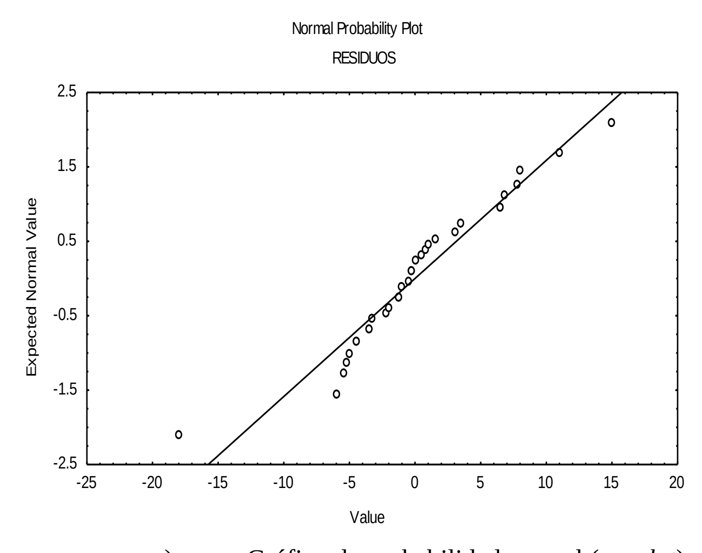
4. Los errores e_{ij} tienen distribución normal (o están normalmente distribuidos). Pruebas de bondad de ajuste, como la de Jí-cuadrado (χ^2), Shapiro-Wilks, Kolmogorov-Smirnov y su modificación conocida como la prueba de Lilliefors, pueden ser utilizadas para examinar esta suposición. Además, la normalidad de los errores puede ser examinada mediante el uso de histogramas, gráficos de cajas de dispersión (*box plot*) y gráficos de probabilidad normal (*normal probability plots*), como se ilustra a continuación:



a) Histograma



b) Diagrama de cajas y alambres (*box plot*)



c) Gráfico de probabilidad normal (*pp-plot*)

Según Nogueira (2007), las suposiciones anteriores tienen por objetivo facilitar la interpretación de los resultados, tornando las técnicas estadísticas más simples, y posibilitando la aplicación de las pruebas de hipótesis. Entretanto, la validez exacta de esas suposiciones es esencialmente teórica, en la práctica, lo que se espera es su validez aproximada, una vez que los procedimientos obtenidos a través de los modelos lineares son razonablemente robustos y se pierde poco sí la validez de las suposiciones fuere apenas aproximada. Cuando esas suposiciones no son satisfechas, esto es, cuando hay desvíos, sus efectos son variados y la gravedad del problema, depende de la situación. De esto modo, se verifica que:

a) La aditividad está asociada a la facilidad de interpretación del modelo, de modo que, si ella es válida, entonces los datos observados serán siempre combinaciones lineares de los efectos estudiados. Esta suposición no es siempre necesaria, sea en la estimación o en las pruebas de hipótesis.

b) La homocedasticidad, o sea, la homogeneidad de varianzas es en la mayoría de veces, el requisito necesario. Cuando se tiene heterogeneidad de varianzas, el método de los mínimos cuadrados no ofrece los mejores estimadores. La prueba de F, los métodos de comparaciones múltiples, y la estimación de los componentes de varianza pueden ser grandemente afectada.

c) En cuanto a la normalidad, es necesaria en las pruebas de hipótesis. En algunas situaciones, no es una suposición crítica, a no ser que vaya acompañada de heterogeneidad de varianzas. Muchas pruebas aplicadas en el análisis de varianza son robustas con relación a la falta de normalidad.

De acuerdo con Macchiavelli (2003), en forma general, la consecuencia del no cumplimiento de los supuestos es que las conclusiones de los análisis realizados pueden no ser válidas (los niveles de error pueden ser diferentes a los establecidos, los errores estándar pueden subestimar o sobreestimar los verdaderos errores poblacionales, los límites de confianza pueden ser incorrectos, etc.). Observando cualquier desvío importante en las suposiciones del análisis de varianza, algunas alternativas pueden ser tomadas:

1. Aplicación de técnicas estadísticas donde no exista la necesidad de suposiciones, tales como: Estadística No Paramétrica, método de los mínimos cuadrados ponderados, método de los residuos específicos.

2. En vez de realizar el análisis estadístico con los datos originales, los cuales no cumplen las suposiciones, se pueden utilizar funciones construidas, de tal modo, que las suposiciones sean cumplidas. Esas funciones construidas son conocidas como: Funciones de transformación de datos (función estabilizadora de la varianza).

3. Uso de Modelos Lineales Generalizados (MLG).

4.2 TRANSFORMACIÓN DE DATOS

La heterogeneidad de varianzas puede ocurrir debido a los tratamientos evaluados, esto es, ciertos tratamientos presentan mayor variabilidad que otros, sin que haya necesariamente una relación entre la media y la varianza, o puede haber sido que si exista esta relación. Esto último significa que la varianza de y es una función de la media, o sea, $E(y) = \mu$ y $V(y) = D^2(\mu)$. En este caso, el procedimiento adoptado se refiere a la transformación de los datos observados a otra escala, antes de realizar el análisis de varianza.

4.2.1 Verificación de los supuestos.

A continuación se presenta un ejemplo tomado del libro Experimentação Agronômica I de la Dra. María Cristina Stolf Nogueira (2007), con el objetivo de ilustrar la metodología para la verificación de los supuestos del Análisis de Varianza y más adelante para mostrar los diferentes tipos de transformaciones que se pueden realizar.

Los siguientes datos se refieren al número de larvas vivas encontradas en un área experimental plantada con arroz *(Oryza sativa* L.), tratada con diferentes insecticidas. El diseño experimental utilizado fue el completamente al azar (DCA).

Tratamiento	Repeticiones				$Y_{i..}$	$\bar{Y}_{i..}$	$\hat{\sigma}^2_{i..}$
	I	II	III	IV			
1	9	12	0	1	22	5.50	35.00
2	4	8	5	1	18	4.50	8.33
3	6	15	6	2	29	7.25	30.25
4	9	6	4	5	24	6.00	4.67
5	27	17	10	10	64	16.00	64.67
6	35	28	2	15	80	20.00	212.67
7	1	0	0	0	1	0.25	0.25
8	10	0	2	1	13	3.25	20.92
9	4	10	15	5	34	8.50	25.67

Modelo matemático-estadístico: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ $\begin{cases} i = 1, 2, \ldots, t \\ j = 1, 2, \ldots, r \end{cases}$

Y_{ij} = número de larvas vivas observado en la ij-ésima unidad experimental.
μ = media general del número de larvas vivas.
τ_i = efecto del i-ésimo insectida.
ε_{ij} = error experimental asociado a la ij-ésima unidad experimental.

ANÁLISIS DE VARIANZA

1. Hipótesis
   $\tau = \tau_i$ (todos los tratamientos producen el mismo efecto)
   $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . , t (al menos uno de los tratamientos produce efectos distintos).

2. Supuestos
   $\varepsilon_{ij} \sim NID(0,\sigma^2)$
   Los errores son independientes y normalmente distribuidos, con media cero y varianza constante (homogeneidad de varianzas).

3. Diagnóstico de los supuestos
   Las posibles violaciones a las suposiciones básicas del modelo estadístico pueden ser investigadas fácilmente examinando los residuos. El residuo de la observación *j* del tratamiento *i* se define mediante:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij}$$

Siendo $\hat{Y}_{ij}$ una estimación de la observación correspondiente, calculada por:

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i$$

$$\hat{Y}_{ij} = \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..})$$

entonces $e_{ij} = Y_{ij} - \overline{Y}_{i.}$, en un diseño completamente al azar (DCA).

Para un diseño aleatorizado por bloques tenemos que:

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$$

$$= \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..})$$

$$= \overline{Y}_{i.} + \overline{Y}_{.j} - \overline{Y}_{..}$$, por lo tanto:

$$e_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$$

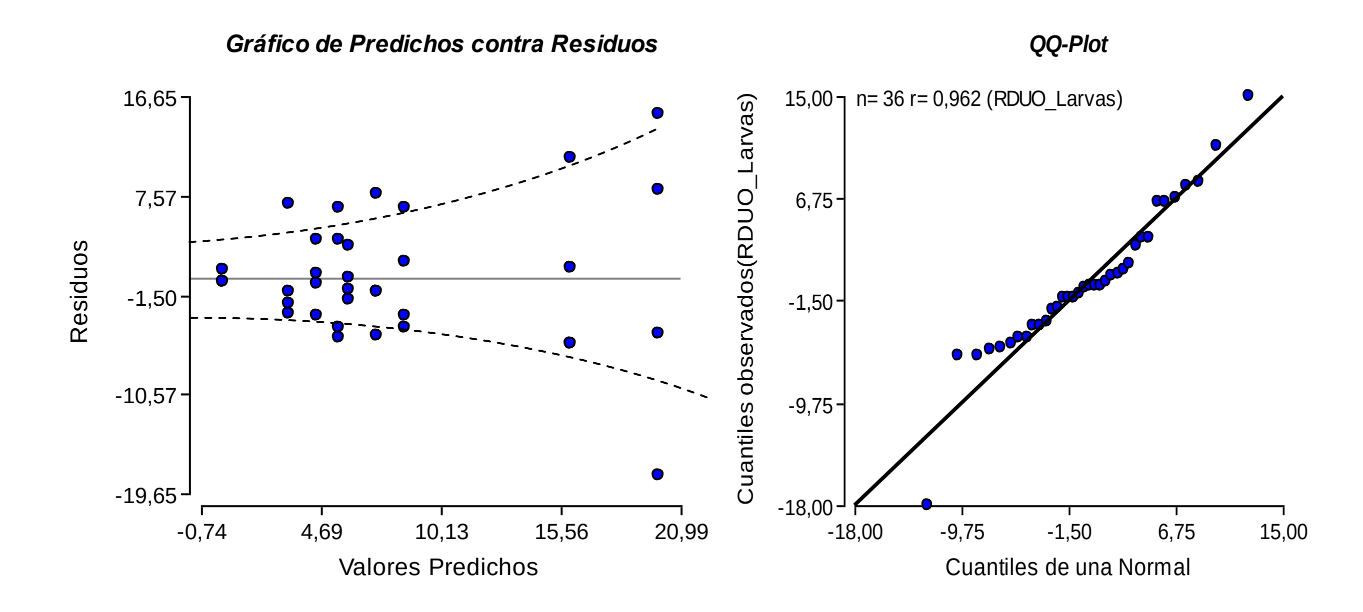
En un diseño cuadrado latino los residuos se obtienen de la siguiente manera:

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk}, \quad \text{o sea:} \qquad e_{ijk} = Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + 2\overline{Y}_{..}$$

4.2.2 Gráfico de residuos contra el valor ajustado $\hat{Y}_{ij}$.

El examen de los residuos debe ser automático en el análisis de varianza. Si el modelo es correcto y las suposiciones se satisfacen, los residuos no deben tener algún patrón, ni deben estar relacionados con alguna otra variable, incluyendo la respuesta y_{ij}. Una comprobación sencilla consiste en graficar los residuos contra los valores ajustados $\hat{Y}_{ij}$ (en el caso de un DCA, recordemos que $\hat{Y}_{ij} = \overline{Y}_{i.}$, el promedio del i-ésimo tratamiento). En esta gráfica no debe revelarse algún patrón obvio.

Un defecto que en ocasiones revela el gráfico es el de una “varianza variable”. Algunas veces la varianza de las observaciones aumenta a medida que la magnitud de las observaciones lo hace. Esto resulta cuando el error es proporcional a la magnitud de la observación. Si éste es el caso, los residuos aumentan a medida que Y_{ij} lo hace. A continuación se presentan el gráfico de predichos contra residuos y el de Quantil-Quantil (QQ-Plot):



Cálculo de los residuos para el ejemplo ilustrativo:

Tratamiento	Repeticiones			
	I	II	III	IV
1	3.50	6.50	-5.50	-4.50
2	- 0.50	3.50	0.50	-3.50
3	-1.25	7.75	-1.25	-5.25
4	3.00	0	-2.00	-1.00
5	11.00	1.00	-6.00	-6.00
6	15.00	8.00	-18.00	-5.00
7	0.75	-0.25	-0.25	-0.25
8	6.75	-3.25	-1.25	-2.25
9	-4.50	1.50	6.50	-3.50

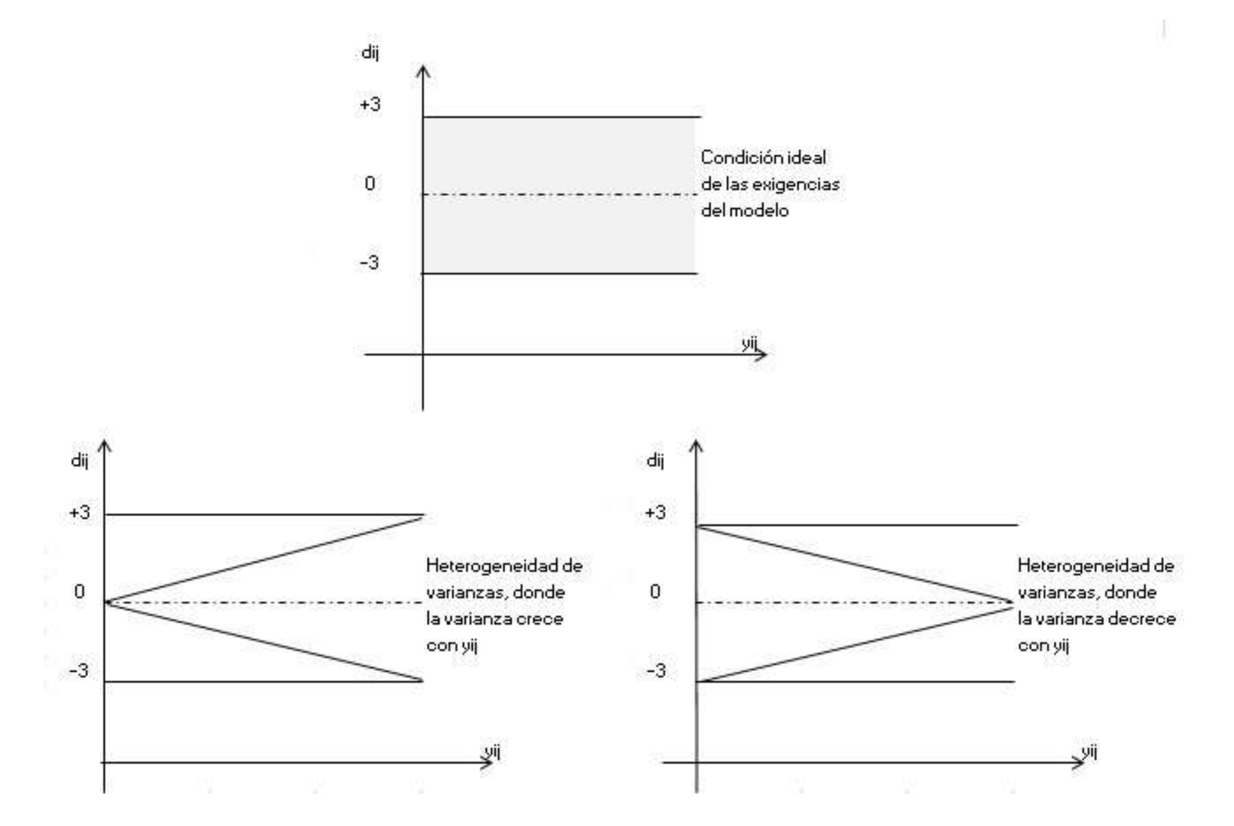
Ejemplo: $e_{11} = 9 – 5.50 = 3.50$, $e_{13} = 0 – 5.50 =$ 3.50,

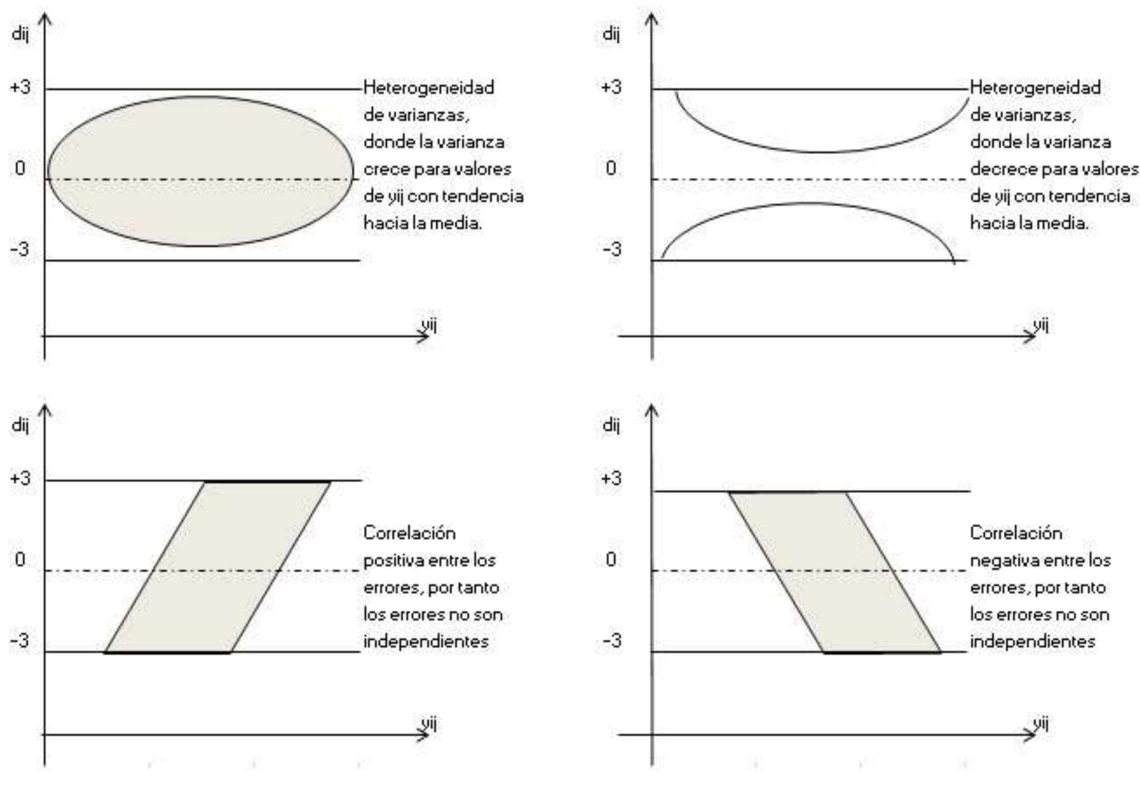
$e_{12} = 12 – 5.50 = 6.50$, $e_{14} = 1 – 5.50 = – 4.50$.

Otra alternativa, según Barbin (2013), es el uso de los residuos estandarizados (d_{ij}), que son calculados usando la siguiente ecuación:

$$d_{ij} = \frac{e_{ij}}{\sqrt{CM_{ee}}} = \frac{e_{ij}}{\sqrt{s^2}},$$

cuando son colocados en un gráfico, contra los valores estimados (y_{ij}), pueden darnos las siguientes orientaciones (patrones de comportamiento):





Fuente: Barbin (2013)

4.2.3 Prueba de Shapiro–Wilk para evaluar el supuesto de normalidad de los errores

En la prueba de normalidad de Shapiro-Wilk, la hipótesis nula que se evalúa es que los errores siguen aproximadamente una distribución normal, y la alternativa es que no siguen esa distribución. Es necesario recordar que si el valor *p* (*Pr* < W) es mayor que el nivel de significancia (en general 0.05), entonces concluimos que no se cuenta con evidencia suficiente para rechazar la hipótesis nula, por lo tanto, los errores siguen una distribución normal.

Shapiro-Wilk

Variable	n	Media	W*	*Pr* < W
RDUO_Rendimiento	20	0.00	0.96	0.7824

En este ejemplo el supuesto de normalidad se acepta Ho , debido a que p = 0.7824 es mayor que el valor de α (0.05).

4.2.4 Pruebas estadísticas para evaluar la homocedasticidad

Además de los gráficos de residuos que frecuentemente se utilizan para diagnosticar la igualdad en las varianzas, se han propuesto algunas pruebas estadísticas. Éstas son pruebas formales para las hipótesis:

Ho : $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_i^2$

Ha : $\sigma_i^2 \neq \sigma_{i'}^2$, para algún $i \neq i'$

- PRUEBA DE HARTLEY O DE F MÁXIMA

Esta prueba consiste en calcular el cociente:

$$\text{Hc} = \frac{\sigma_i^2(\text{máx})}{\sigma_i^2(\min)}, \text{i} = 1, 2, \ldots, \text{t}$$

siendo:

$$\sigma_i^2 = \frac{1}{r-1}\left(\sum_{j=1}^{r} Y_{ij}^2 - \frac{Y_{i.}^2}{r}\right)$$, siendo

- t = número de tratamiento y
- r= número de repeticiones

Regla de decisión:
Si Hc ≥ $H_{(t,\, r-1,\, \alpha)}$, se rechaza Ho, con un nivel de significancia α. Siendo $H_{(t,\, r-1,\, \alpha)}$ obtenido de la Tabla 5 del Apéndice. Esta tabla fue generada por Pearson & Hartley en 1956.

Para el ejemplo ilustrativo, se tiene que en el tratamiento con el insecticida 6 se obtuvo la mayor varianza estimada:

$$\sigma_i^2 = \frac{1}{4-1}\left[(35^2 + 28^2 + 2^2 + 15^2) - \frac{80^2}{4}\right] = 212.67 \text{ u}^2$$

Por otra parte, los datos observados en el tratamiento 7 presentaron el menor valor de $\sigma_i^2 = 0.25 \text{ u}^2$. Por tanto,

$$\text{Hc} = \frac{212.67}{0.25} = 850.68 \quad \text{y} \quad H_{(9,3,0.05)} = 93.9 \text{ (vea la Tabla 1)}$$

Como Hc ≥ H, se rechaza la hipótesis nula y se concluye que existen evidencias de ocurrencia de heterocedasticidad (o sea, heterogeneidad de varianzas).

- PRUEBA DE BARTLETT (Snedecor e Cochran, 1983)

Consiste en calcular una estadística cuya distribución muestral es, aproximadamente Ji cuadrada con k-1 grados de libertad, cuando las k muestras aleatorias (o grupos) provienen de poblaciones normales e independientes. La estadística de prueba es:

$$\chi_o^2 = 2.3026 \times \frac{q}{c}$$

$$q = (N-k) \times \log_{10} S_p^2 - \sum_{i=1}^{k} (n_i - 1) \log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}(n_i - 1)^{-1} - (N-k)^{-1}\right]$$

$$S_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1) \times S_i^2}{N - k}$$, siendo S_i^2 es la varianza muestral de la i-ésima población.

Regla de decisión:

Se rechaza Ho para valores grandes de χ_o^2; en otras palabras, se rechaza Ho, sólo si: $\chi_o^2 > \chi^2_{(k-1,\alpha)}$. El valor de $\chi^2_{(k-1,\alpha)}$ se obtiene de la Tabla 6 del Apéndice.

Nota: Varios estudios indican que la prueba de Bartlett es muy sensible a la suposición de normalidad y no debe ser aplicada cuando exista alguna duda en cuanto a esta suposición (Montgomery, 2004).

Retomando al ejemplo ilustrativo que se ha venido trabajando, tenemos que:

N = total de observaciones
k = número de grupos (o tratamientos)
Cuando todas las estimativas S_i^2 son obtenidas con el mismo número de observaciones tenemos que S_p^2 se obtiene de la siguiente manera:

$$S_p^2 = (4-1) \times \frac{(35 + 8.33 + 30.25 + \ldots + 25.67)}{(36-9)} = 44.71$$

$$q = (36-9) \times \log_{10}(44.71) - (4-1) \times \left[\log_{10}(35) + \log_{10}(8.33) + \ldots + \log_{10}(25.67)\right] = 11.92$$

$$c = 1 + \frac{1}{3(9-1)} \times \left[\left(\frac{1}{(4-1)} \times 9\right) - \frac{1}{(36-9)}\right] = 1.12$$

$$\chi_o^2 = 2.3026 \times \frac{11.92}{1.12} = 24.51$$

$$\chi^2_{(8,0.05)} = 15.51$$

Como $\chi_o^2 > \chi^2_{(8,0.05)}$, se rechaza Ho, concluyendo que las varianzas son heterogéneas.

4.2.5 Algunas transformaciones comúnmente utilizadas

El proceso de obtención de la transformación de datos depende del objetivo que se desea alcanzar con esa transformación. Es sabido que difícilmente una transformación de datos irá atender a todos los objetivos, aunque no sea raro, se busca homogeneidad de varianzas y se consigue junto con ella una mejor aproximación normal.

En la práctica, raramente se realizaba la verificación de las condiciones exigidas por el modelo matemático, debido a que, datos de peso y altura (variables cuantitativas continuas), generalmente tienen distribución normal. A continuación se describen los cuatro principales tipos de transformaciones comúnmente utilizados:

- Raíz cuadrada.

Es una transformación usualmente utilizada para datos con varianzas que cambian proporcionalmente con la media, frecuentemente cuando la variable observada Y se refiere a datos de conteo de insectos u otros organismos, permitiendo suponer que Y tiene distribución de Poisson.

Variables de este tipo deben transformarse mediante la raíz cuadrada:

$$y_{ij}* = \sqrt{y_{ij}} \quad \text{o} \quad \text{bien} \quad y_{ij}* = \sqrt{y_{ij} + \lambda}$$

siendo λ = 0, ½, 3/8, etc.

Gomes (2000) indica que cuando se incluyen valores de y_{ij} inferiores a 15, conviene tomar λ = 3/8 (que es lo que comúnmente ocurre con experimentos con insectos en laboratorio). Cuando los valores de y_{ij} son mayores de 15, se puede tomar $\lambda = 0$.

- Angular

Otro caso a considerar es el que trata de porcentajes p= (x/n) × 100, relativas a n observaciones por parcela. Los porcentajes deben estar basados en un denominador común (por ejemplo, porcentaje de germinación calculado a partir de 50 semillas bajo distintos tratamientos). En tales condiciones, los datos tienen, en general, distribución binomial y la transformación indicada es:

$$y_{ij} = \text{arcsen } (y_{ij})^{1/2} \quad \text{o} \quad \text{sea} \quad y_{ij}* = \text{sen}^{-1} (y_{ij})^{1/2}$$

Gomes (2000) recomienda que se efectúe este tipo de transformación cuando los porcentajes sean menores de 15% o excedan 85%. Así, si todos los datos se encuentran en el intervalo [15%, 85%], la transformación no es necesaria. Además, en el caso en que x=0, el valor de 0/n debe ser substituido por (1/4n), y en el caso de x=n, el valor de n/n será substituido por 1–(1/4n).

Los datos transformados se expresan en grados (y no en radianes), por ejemplo, si y_{ij} = 5%, y_{ij}*= arcsen$(0.05)^{1/2}$ =12.92.

- Logarítmica

Cuando se verifica una proporcionalidad entre medias y desviaciones estándar, se puede usar la transformación:

$$y_{ij} = \log (y_{ij}) \quad \text{o} \quad \text{sea} \quad y_{ij}* = \log (y_{ij} + 1)$$

Macchiavelli (2003) indica que esta transformación se utiliza para datos que exhiben efectos multiplicativos (una forma de falta de aditividad) o cuando las varianzas son proporcionales al cuadrado de las medias. Gomes (2000) cita que, la transformación logarítmica puede utilizarse también, cuando los datos siguen una distribución log-normal.

- Recíproca

Surge en las situaciones en que V (y_{ij}) $\propto$ (es proporcional) a μ^4. Entonces:

$$y_{ij}* = \frac{1}{y_{ij}}$$

La utilización de esta transformación es adecuada cuando *y* está relacionada con el tiempo de sobrevivencia, luego 1/*y* puede ser considerada como la tasa de mortalidad, o cuando *y* está relacionada con el tiempo transcurrido hasta un acontecimiento o un número de ocurrencias para una unidad de tiempo.

4.2.6 Selección de la transformación.

Montgomery (2004) indica que si el experimentador conoce la relación entre la varianza y la media de las observaciones, puede utilizar esa información como guía al seleccionar la forma de la transformación apropiada.

Sea E(y)=μ, la media de los valores de y. Y Suponiendo que la varianza V(y)= σ^2_y, es proporcional a alguna potencia de la media de y, tal que: $\sigma^2_y \propto \mu^\alpha$, se desea determinar la transformación de y que produzca una varianza constante.

La transformación recomendable frecuentemente puede ser obtenida con el auxilio de la ecuación de regresión: $V(y) = \theta \times \hat{\mu}^b$, en donde θ es una constante de proporcionalidad. Al aplicar logaritmo a ambos lados de la anterior ecuación de regresión, se obtiene la ecuación (1):

$$Ln(\hat{\sigma}_i^2) = \hat{\alpha} + \hat{b} \times Ln(\bar{y}_{i.}) \qquad i = 1, 2, ..., t$$

La metodología se describe a continuación:

1. Se deben obtener las estimativas de la media y varianza de los tratamientos.
2. Luego se debe realizar un análisis de regresión entre los valores de $\text{Ln}(\hat{\sigma}_\text{i}^2)$ en función de los valores de $\text{Ln}(\bar{y}_{\text{i.}})$ utilizando la ecuación (1).

Para el ejemplo ilustrativo que hemos venido trabajando, tenemos los siguientes resultados:

Insecticida	$\bar{y}_{i.}$	$x_i = Ln(\bar{y}_{i.})$	$\hat{\sigma}_i^2$	$y_i = Ln(\hat{\sigma}_i^2)$
1	5.50	1.7047	35.00	3.5553
2	4.50	1.5041	8.33	2.12
3	7.25	1.9810	30.25	3.4095
4	6.00	1.7918	4.67	1.5412
5	16.00	2.7726	64.67	4.1693
6	20.00	2.9957	212.67	5.3597
7	0.25	-1.3863	0.25	-1.3863
8	3.25	1.1787	20.92	3.0407
9	8.50	2.1401	25.67	3.2453
Total		14.6824		25.0547

Con esta información obtenemos la estimativa del coeficiente angular de la regresión lineal simple:

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = 1.3918$$

Para evaluar la hipótesis Ho: β=0 contra Ha: β≠0, se realiza el análisis de varianza para la regresión lineal, y los resultados se presentan en el siguiente cuadro:

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
Regresión	1	24.99375	24.99375	41.6306*	5.59
Resíduos	7	4.20259	0.60037		
Total	8	29.19634			

Coeficiente de determinación (R^2) = 0.8561

Por el resultado obtenido, con un 5% de significancia, no se encontró evidencia suficiente para rechazar la hipótesis nula, por lo tanto se concluye que existe relación entre la media y la varianza de los tratamientos incluidos en el experimento del ejemplo ilustrativo.

La selección de la transformación adecuada depende de la estimativa del coeficiente angular de la regresión ($\hat{b}$). En la siguiente tabla se presentan las principales transformaciones a ser adoptadas, en función del valor del coeficiente $\hat{b}$

Relación entre V(y)= σ^2 y μ	$\hat{b}$	$\gamma = 1 - \left(\frac{\hat{b}}{2}\right)$	Transformación
$\sigma^2 \propto$ Constante	0	1	Ninguna
$\sigma^2 \propto \mu$	1	½	$\sqrt{y+\lambda}$
$\sigma^2 \propto \mu$	2	0	$\log(y+\lambda)$
$\sigma^2 \propto \mu$	3	–½	$1/\sqrt{y+\lambda}$
$\sigma^2 \propto \mu$	4	–1	$1/(y+\lambda)$

Observación: λ puede asumir valores tales como: 0, 3/8, ½, 1, 2, . . . , con el objetivo de corregir valores nulos o negativos. La transformación a utilizar en nuestro ejemplo sería:

$$y_{ij}^* = y_{ij}^{\left(1-\frac{\hat{b}}{2}\right)} \qquad \therefore \quad 1-\frac{\hat{b}}{2} = 1-\frac{1.3918}{2} = 0.3041 \qquad y_{ij}^* = y_{ij}^{(0.3041)}$$

Al utilizar la tabla anterior, se puede verificar que la transformación adecuada es "raíz cuadrada", considerando $\hat{b} = 1.3916 \cong 1$.

4.3 PROGRAMA EN SAS PARA EL ANÁLISIS DE LOS SUPUESTOS DE NORMALIDAD Y DE HOMOGENEIDAD DE VARIANZAS

```
OPTIONS nodate nonumber;
DATA supos; /*evaluacion de los supuestos de normalidad y homocedasticidad*/
INPUT insec larvas;
CARDS;
1 9
2 4
3 6
4 9
5 27
6 35
7 1
8 10
9 4
1 12
2 8
3 15
4 6
5 17
6 28
7 0
8 0
9 10
1 0
2 5
3 6
4 4
5 10
6 2
7 0
8 2
9 15
1 1
2 1
3 2
4 5
5 10
6 15
7 0
8 1
9 5
;
PROC glm;
TITLE "Prueba de normalidad";
CLASS insec;
MODEL larvas=insec/SS1;
output out=res  r=rlar p=pred;
RUN;
```

Otra alternativa, es digitando luego del conjunto de datos, esta parte del programa:

```
PROC GLM DATA=supos;
CLASS insec;
MODEL larvas=insec;
OUTPUT OUT=resi P =preditos R=resid STUDENT=res_pad; /*se guardan los residuos y los residuos estudentizados en el archivo RESI*/
/*obtención de las pruebas de normalidad y gráfico de probabilidad normal*/
PROC UNIVARIATE DATA=resi NORMAL PLOT; /*con la opción NORMAL se muestran las pruebas de normalidad y con PLOT el gráfico de probabilidad NORMAL*/
VAR resid;
QQPLOT resid; /*se presenta el gráfico QQPLOT*/
/*obtención de las pruebas de homocedasticidad de los residuos agrupados por insecticida*/
PROC GLM DATA=resi;
CLASS insec;
MODEL larvas=insec;
MEANS insec/HOVTEST=BARTLETT; /*se obtienen las pruebas de BARTLETT*/
MEANS insec/HOVTEST welch; /*se obtienen las pruebas de LEVENE y WELCH*/
RUN;
title 'Gráfico dos resíduos padronizados - independência';
proc plot;  plot res_pad*preditos;
run;
```

```
PROC print data=res; RUN;
PROC univariate plot normal data=res;
var rlar;
RUN;

TITLE "Verificación de la homocedasticidad usando la prueba de Hartley";
proc means mean var CV;
 by insec;
 var larvas;
RUN;

PROC SUMMARY NWAY;/* Calcula y guarda varianza estimada */
TITLE “Prueba de Bartlett para contrastar igualdad de varianzas”;
          CLASS insec;/* el número de observaciones para cada nivel de trat*/
    VAR larvas;
    OUTPUT OUT= salida  VAR=VARIANCE N=NUM;
    PROC PRINT;
RUN;

    DATA _NULL_;
    SET SALIDA END=EOF;
    LOGVARI=LOG(VARIANCE);
    N=NUM-1;
    SLOGVAR+LOGVARI*N;
    TOTN+N;
    NVAR=N*VARIANCE;
    SNVAR+NVAR;
    A+1;
    SFRACT+1/N;
    IF EOF THEN DO;
    M=TOTN*LOG(SNVAR/TOTN)-SLOGVAR;
    C=1+(1/(3*(A-1)))*(SFRACT-1/TOTN);
    CHISQ=M/C;
    PROBCHI=PROBCHI(CHISQ,(A-1));
    ALPHA=1-PROBCHI;
    FILE PRINT;
    PUT  “Test de Bartlett para contrastar igualdad de varianzas”;
    PUT   “ ”;
    PUT  “CHI-CUADRADO =” CHISQ ”  ALPHA =” ALPHA ”. ”;
    END;
    RUN;
```

4.4 EJERCICIOS PROPUESTOS

1. Efectúe el análisis de varianza para la variable raíz cuadrada del número de larvas vivas. Los datos (ya transformados) se presentan en el cuadro siguiente:

TRAT.	Repeticiones				$Y_{i..}$	$\bar{Y}_{i.}$
	I	II	III	IV		
1	3.00	3.46	0	1.00	7.46	1.87
2	2.00	2.83	2.24	1.00	8.07	2.02
3	2.45	3.87	2.45	1.41	10.18	2.55
4	3.00	2.45	2.00	2.24	9.69	2.42
5	5.20	4.12	3.16	3.16	15.64	3.91
6	5.92	5.29	1.41	3.87	16.49	4.12
7	1.00	0	0	0	1.00	0.25
8	3.16	0	1.41	1.00	5.57	1.39
9	2.00	3.16	3.87	2.24	11.27	2.89

2. Navarro (1996) evaluó el efecto de aceites y detergentes para el control de áfidos *(Homóptera: Aphididae)* en el cultivo del brócoli *(Brassica oleracea* var. Itálica Plenk), en el municipio de Patzicía, departamento de Chimaltenango. A continuación se presentan los datos de campo del conteo del número de áfidos por parcela neta, 7 días después de la segunda aplicación de los tratamientos. El diseño experimental utilizado fue el de bloques completos al azar (vea el capítulo 5).

Tratamientos	Repeticiones		
	I	II	III
Saf-T-Side (Aceite parafinado)	49	26	35
Triona o Medopaz (Aceite mineral de uso agrícola)	19	49	27
Carrier (Aceite vegetal de uso agrícola)	32	43	42
Olmeca (Aceite vegetal de uso doméstico)	13	13	17
Bioambar (Detergente biodegradable)	41	40	51
Glicerina (Detergente en formulación líquida)	123	10	60
Sulfonato de potasio (Detergente en formulación semisólida)	3	70	87
ACT-92 (Carbonato de sodio)	26	70	14
Naled (Testigo químico, insecticida organofosforado)	61	43	4
Testigo absoluto (Agua)	43	37	62

a) Evalúe los supuestos del Análisis de Varianza.
b) En caso de ser necesario, indique y demuestre cuál es la transformación más adecuada.

3. Un experimento reportado por Macedo (1970) y citado por Banzatto y Kronka (2011) fue realizado con el objetivo de evaluar 5 tratamientos para el control del pulgón (*Aphis gossypii* Glover) en el cultivo del pepino (*Cucumis sativus*), fueran utilizadas 6 repeticiones y el diseño adoptado fue el de completamente al azar. Los datos obtenidos referentes al número de pulgones colectados 36 horas después de la fumigación, son presentados en el cuadro siguiente:

Tratamientos	Repeticiones						s^2
	1	2	3	4	5	6	
A	2370	1687	2592	2283	2910	3020	233750
B	1282	1527	871	1025	825	920	75559
C	562	321	636	317	485	842	40126
D	173	127	132	150	129	227	1502
E	193	71	82	62	96	44	2792

a) Verifique el cumpliento de los supuestos del modelo estadístico matemático: normalidad de los residuos y homogeneidad de las varianzas.
b) Indique el tipo de heterogeneidad. Puede utilizar gráficos de los residuos.
c) Justifique el tipo de transformación más adecuada y aplíquela.
d) Con los datos transformados verifique de nuevo los supuestos del modelo estadístico matemático.
e) Realice el ANOVA con los datos transformados.
f) En caso de ser necesario, realice la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey.

4. El porcentaje de humedad relativa (HR) es determinante para el ataque (%) de hongos en semillas. Para evaluar la susceptibilidad de las semillas de maní al ataque de un hongo, Di Rienzo *et al*., citan un experimento que se realizó en cámaras de cría con tres porcentajes de HR: 70%, 80% y 90%. Cinco observaciones fueron tomadas para cada porcentaje de HR, registrándose el porcentaje de semillas atacadas en un grupo de cien semillas (unidad experimental). Los resultados se presentan a continuación:

Porcentaje de HR	Observaciones (Porcentaje de semillas infectadas)				
70	7	6	9	5	9
80	12	15	17	18	20
90	14	16	18	21	15

a) Verifique el cumpliento de los supuestos del modelo estadístico matemático: normalidad de los residuos y homogeneidad de las varianzas.
b) Indique el tipo de heterogeneidad. Puede utilizar gráficos de los residuos.
c) Realice el ANOVA.

5. Revise tres trabajos de investigación en los que la variable de respuesta medida sea una variable cuantitativa discreta, analice los datos y verifique si los supuestos del modelo estadístico matemático fueron cumplidos. Indique qué tipo de transformación es la más adecuada para cada caso, justificando su selección.

CAPÍTULO 5
DISEÑO EN BLOQUES COMPLETOS AL AZAR

5.1 INTRODUCCIÓN

El diseño en bloques completos al azar (DBA) toma en cuenta los tres principios básicos de la experimentación: repetición, aleatorización y control local. En este diseño las unidades experimentales se distribuyen en grupos homogéneos. Cada uno de estos grupos es llamado: bloque. El número de unidades experimentales dentro de cada bloque es igual al número de tratamientos incluidos en el experimento. Un caso particular de diseño de bloques es el que aparece relacionado con la prueba de t para muestras pareadas, aunque el número de tratamientos es sólo dos.

Los tratamientos son distribuidos en las unidades experimentales dentro de cada bloque aleatoriamente, así, cada bloque irá a constituir una repetición. Este tipo de experimento es seleccionado cuando se tienen dudas acerca de la homogeneidad del ambiente o cuando, por experiencia, se sabe de su heterogeneidad.

5.1.1 Criterios de bloqueo

Este diseño es conveniente cuando se logra determinar un gradiente de variabilidad en un sentido, que esté influyendo sobre los tratamientos, por ejemplo: grado de inclinación del terreno donde se realizará el experimento, dirección del viento, gradiente de temperatura, gradiente de fertilidad, de luminosidad, etc. Los bloques se construyen perpendiculares a la dirección del gradiente de variabilidad.

5.1.2 Aleatorización

Se aleatorizan las tratamientos dentro de cada bloque. Debe considerarse que la aleatorización se realizará de forma independiente para cada bloque.

Ejemplo

Suponiendo que se tiene un experimento que incluye un factor con 5 niveles (denotados con las letras A, B, C, D y E), distribuidos en 4 bloques, y considerando que en cada bloque los niveles del factor fueron totalmente aleatorizados, el croquis de campo quedaría de la siguiente forma:

A	C	D	E	B	BLOQUE I
D	A	B	E	A	BLOQUE II
D	B	A	C	E	BLOQUE III
E	D	B	A	C	BLOQUE IV

El proceso de aleatorización puede ser simplificado, utilizando el procedimiento PROC PLAN del programa SAS, conforme se presenta a continuación, para el ejemplo anterior, 5 tratamientos en 4 bloques.

****

Programa

```
Title "Aleatorización de los tratamientos en un DBA";
PROC Plan seed=1820403;
Factors bloques=4 parcelas=5 ORDERED;
Treatments Trat=5 RANDOM;
Output out=croquis Trat cvals= ("A" "B" "C" "D" "E");
RUN;
PROC Print DATA=croquis;
RUN;
```

****

Salida

Aleatorización de los tratamientos en un DBA

The PLAN Procedure

Plot Factors			
Factor	Select	Levels	Order
Bloques	4	4	Random
Parcelas	5	5	Ordered

Treatment Factors			
Factor	Select	Levels	Order
Trat	5	5	Random

bloques	parcelas					Trat				
2	1	2	3	4	5	3	1	2	4	5
4	1	2	3	4	5	2	1	5	3	4
1	1	2	3	4	5	1	2	5	4	3
3	1	2	3	4	5	5	1	4	2	3

Para realizar el proceso de aleatorización en otros diseños experimentales utilizando PROC Plan de SAS, puede consultar el libro “Experimentação Agronômica I” de la Dra. Maria Cristina Stolf Nogueira (2007). En el programa R también es posible realizar la aleatorización de los tratamientos. Para ampliar esta parte puede comunicarse por email con los autores del presente texto (**Ezequiel López o Byron González**).

El croquis de campo para este ejemplo se presenta a continuación.

Obs	Bloques	Parcelas	Trat	Obs	Bloques	Parcelas	Trat
1	2	1	C	11	1	1	A
2	2	2	A	12	1	2	B
3	2	3	B	13	1	3	E
4	2	4	D	14	1	4	D
5	2	5	E	15	1	5	C
6	4	1	B	16	3	1	E
7	4	2	A	17	3	2	A
8	4	3	E	18	3	3	D
9	4	4	C	19	3	4	B
10	4	5	D	20	3	5	C

5.2 ANÁLISIS ESTADÍSTICO

Sea t el número de niveles del factor A (tratamientos) distribuidos en r bloques. La notación adoptada para representar los valores de la variable de respuesta es dada en la tabla siguiente:

Tratamientos	Repeticiones					$Y_{i.}$
	1	2	3	. . .	R	
1	Y_{11}	Y_{12}	Y_{13}	. . .	Y_{1r}	$Y_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	. . .	Y_{2r}	$Y_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	. . .	Y_{3r}	$Y_{3.}$
.	.	.	.	. . .	.	.
.	.	.	.	. . .	.	.
T	Y_{t1}	Y_{t2}	Y_{t3}	. . .	Y_{tr}	$Y_{t.}$
$Y_{.j}$	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	. . .	$Y_{.r}$	$Y_{..}$

Siendo que:

$Y_{.j} = \sum_{i=1}^{t} Y_{ij}$ es el total obtenido en el j-ésimo bloque o repetición

$Y_{i.} = \sum_{j=1}^{r} Y_{ij}$ es el total obtenido en el i-ésimo tratamiento

$Y_{..} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}$ es el total general o gran total

$\overline{Y}_{..} = \frac{Y_{..}}{tr}$ es la media general $\overline{Y}_{i.} = \frac{Y_{i.}}{r}$ es la media del i-ésimo tratamiento

5.2.1 Hipótesis a evaluar

Ho : $\forall_i$ i = 1,2, . . . / $\tau_i = \tau$
Ha: $\exists_i$ i = 1,2, . . . / $\tau_i \neq \tau$

5.2.2 Modelo estadístico

El modelo asociado a este diseño experimental se muestra a continuación:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \ldots, t \\ j = 1, 2, 3, \ldots, r \end{cases}$$

Siendo:

Y_{ij}	=	variable de respuesta observada o medida en el i-ésimo tratamiento y el j-ésimo bloque.
μ	=	media general de la variable de respuesta
τ_i	=	efecto del i-ésimo tratamiento
β_j	=	efecto del j-ésimo bloque
ε_{ij}	=	error asociado a la ij-ésima unidad experimental.

5.2.3 Supuestos

$\varepsilon_{ij} \sim NID(0, \sigma^2)$
No existe interacción entre bloque y tratamiento, lo que significa que un tratamiento no debe modificar su acción (o efecto) por estar en uno u otro bloque.

5.2.4 Análisis de Varianza

FV	GL	SC	CM	Valor de F
Bloques	r- 1	$\sum_{j=1}^{r} \frac{Y_{.j}^2}{t} - \frac{Y_{..}^2}{tr}$		
Tratamientos	t – 1	$\sum_{i=1}^{t} \frac{Y_{i.}^2}{r} - \frac{Y_{..}^2}{tr}$	SC_{trat} / gl_{trat}	CM_{trat}/CM_{ee}
Error experimental	(t -1) (r -1)	$SC_{total} - (SC_{trat} + SC_{bloque})$	SC_{ee} / gl_{ee}	
Total	tr – 1	$\sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij}^2 - \frac{Y_{..}^2}{tr}$		

$$CV(\%) = \frac{\sqrt{CME}}{\overline{Y}..} \times 100$$

Regla de Decisión

Rechazar H_o. Sí el valor de $F \geq$ F crítica $_{(gl\ trat;\ gl\ error;\ \alpha)}$
No Rechazar H_o. Sí el valor de $F <$ F crítica $_{(gl\ trat;\ gl\ error;\ \alpha)}$

5.2.5 Ejemplo de aplicación

En un experimento se evaluó la aplicación de productos químicos para el control de nematodos. Fueron utilizados los siguientes tratamientos:

A. Testigo absoluto (sin aplicación).
B. Oxamyl 1.5 lt (forma de aplicación: foliar)
C. Oxamyl 1.5 lt (forma de aplicación: al suelo)
D. Oxamyl 2.0 lt (forma de aplicación: foliar)
E. Carbofuran 15 gr. (aplicado al suelo)
F. Oxamyl 2.0 lt (forma de aplicación: al suelo)

Los tratamientos fueron analizados en un diseño bloques al azar con cinco repeticiones. La variable de respuesta medida fue el número de nematodos vivos por unidad experimental. Los datos obtenidos se presentan en la siguiente tabla:

Nematicidas	Bloques					$Y_{i.}$	$\overline{Y}_{i.}$
	I	II	III	IV	V		
A	307	371	379	360	339	1756	351.2
B	187	192	320	243	296	1238	247.6
C	277	328	363	195	344	1507	301.4
D	115	235	248	267	256	1121	224.2
E	173	267	251	254	200	1145	229.0
F	195	131	171	253	253	1003	200.6
$Y_{.j}$	1254	1524	1732	1572	1688	7770	259

Se le solicita realizar el análisis de varianza y emitir las respectivas conclusiones.

Solución:

$$SCtrat = \frac{(1756)^2 + (1238)^2 + (1507)^2 + (1121)^2 + (1145)^2 + (1003)^2}{5} - \frac{(7770)^2}{(5)(6)} = 79,750.8$$

$$SCbloques = \frac{(1254)^2 + (1524)^2 + (1732)^2 + (1572)^2 + (1688)^2}{6} - \frac{(7770)^2}{(5)(6)} = 23,477.33$$

$$SCtotal = (307^2 + 371^2 + 379^2 + \ldots + 253^2) - \frac{(7770)^2}{(5)(6)} = 148,922$$

Resumen del análisis de varianza

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	Valor crítico de F
Nematicidas	5	79750.80	15950.16	6.98*	2.71
Bloques	4	23477.33	5869.33		
Error Experimental	20	45693.87	2284.69		
Total	29	148922.00			

$$CV(\%) = \frac{\sqrt{2284.69}}{259} \times 100 = 18.45\%$$

Conclusión: Los nematicidas evaluados producen diferente efecto en el control de nematodos.

5.3 EJERCICIOS PROPUESTOS, DBA

1. Los datos siguientes fueron obtenidos de un experimento realizado en el Departamento de Botánica de la ESALQ/USP en 1988. Fue utilizado un diseño en bloques completos al azar, con 4 bloques y 5 tratamientos (T_1 = control, T_2 = Giberelina 50 ppm, T_3 = Ácido naftalenoacético 100 ppm, T_4 = Chlormequat 1,500 ppm y T_5=Daminozide 3,000 ppm). La variable de respuesta fue rendimiento de frutos (gramos por parcela) de la variedad de fresa (*Fragaria vesca*) Campinas. Cada unidad experimental estuvo formada por dos líneas de 1.5 × 2.0 m, con 10 plantas. Los resultados se presentan a continuación:

Tratamientos	Bloques				
	I	II	III	IV	$Y_{i.}$
T_1	374.68	459.15	306.79	350.32	1490.94
T_2	524.67	281.95	294.41	405.91	1506.94
T_3	329.10	258.81	294.04	443.72	1325.67
T_4	164.02	378.15	297.22	275.75	1115.14
T_5	451.57	417.70	348.12	351.56	1568.95

Realice el ANOVA y en caso de ser necesario una prueba post-ANOVA.

2. En un experimento de riegos en el cultivo de algodón (*Gossypium hirsutum* L.) se evaluaron los siguientes tratamientos que están expresados en metros cúbicos de agua absorbidos por hectárea: T_1=5400, T_2=4800, T_3=4200 y T_4=3600. El experimento se condujo en parcelas de 300 m^2 de área útil y los resultados están expresados en kilogramos por unidad experimental.

Bloques	Tratamientos				
	T_1	T_2	T_3	T_4	$Y_{.j}$
I	68	73	53	50	244
II	86	90	62	62	300
III	68	71	46	50	235

a) Plantee el modelo y defina sus componentes
b) Defina las hipótesis y construya el cuadro de ANOVA. Utilice α=0.05
c) Efectúe la prueba de Tukey para comparar las medias de los tratamientos.

3. Dentro de los estándares de calidad del fruto de mango (*Mangifera indica* L.) para que sea exportable como un fruto de primera debe de cumplir con lo siguiente: poseer por lo menos un 20% de coloración rojiza, sin ningún daño mecánico y que no se encuentre manchado; en la región de la costa sur está el problema de la falta de coloración rojiza del fruto, por lo cual se llevó a cabo una investigación en el cultivo de mango (*Mangifera índica* L. Var. *Tommy atkins)* en la cual se evaluó el efecto de la fertilización foliar en la época de floración y crecimiento del fruto sobre la calidad del fruto.

Una de las variables de respuesta evaluadas fue el número de frutos con coloración roja. Se utilizó el diseño de bloques al azar debido a que existe una gradiente de variabilidad en la plantación. Los resultados obtenidos se presentan a continuación:

CUADRO 1: Tratamientos evaluados

Tratamiento	Código
Potasio + Magnesio + Balanceador + cobre	Mezcla 1
Mono-di-tri-polisacaridos + Ácidos urónicos + MgO, CaO +B, Zn, Co.	Mezcla 2
Nitrato de Potasio + Aminoácidos +Ácido fólico + Nutrientes	Mezcla 3
Testigo	Sin mezcla

CUADRO 2 Número de frutos con coloración roja en parcelas donde se realizó una poda de aclareo en la etapa de floración

Tratamiento	Repetición			
	I	II	III	IV
Mezcla 1	637	572	858	732
Mezcla 2	731	603	530	797
Mezcla 3	687	561	489	582
Sin mezcla	547	640	644	586

4. En Córdoba, Argentina se realizó un ensayo para evaluar el rendimiento en kg de materia seca por hectárea de una forrajera megatérmica con distintos aportes de N_2 en forma de urea. Las dosis de urea probadas fueron: 0 (control), 75, 150, 225 y 300 kg.ha^{-1}. El ensayo se realizó en distintas zonas, en las que por razones edáficas y climáticas se podían prever rendimientos diferentes. Las zonas en este caso actuaron como bloques. Los resultados obtenidos por tratamiento y bloque se presentan a continuación:

Urea (kg/ha)	Bloque			
	I	II	III	IV
0 (control)	2010	1832	2170	1879
75	2915	2175	2610	2294
150	3049	2908	2964	2971
225	3199	3235	3003	2937
300	3381	3270	3129	3171

a) Plantee el modelo y defina sus componentes
b) Defina la hipótesis y construya el cuadro de ANOVA. Utilice α=0.05
c) Efectúe la prueba de Duncan para comparar las medias de los tratamientos.

5. Peña Hernández (1997) realizó una investigación en tres localidades del departamento de El Progreso (Guatemala), con el objetivo de evaluar el rendimento de catorce líneas de frijol Tepary, procedentes de Hermosillo, Estado de Sonora, México. El diseño experimental utilizado en cada local fue bloques al azar, con cuatro repeticiones. La unidad experimental fue de dimensiones 9,40 m × 48 m, con distanciamiento de 0.30 m entre por plantas y 0.40 m entre surcos. La variable de resposta medida fue el rendimiento de granos de frijol (kg.ha^{-1}). Realice el análisis de varianza para cada local con alfa =5%, y en caso de ser necesario, aplique una prueba de comparación múltiple de medias. Compare los resultados obtenidos.

El Jícaro

	Bloques			
Variedades	1	2	3	4
L-18	405.56	526.35	245.10	458.81
L-242-43	755.35	555.96	615.70	533.27
L-242-7	348.87	359.16	467.32	465.94
L-39	232.4	341.65	209.17	462.48
L-246-19	400.48	526.23	284.02	389.48
L-242-22	373.49	867.97	589.55	686.75
L-242-9	265.69	266.96	443.40	477.13
L-246-9	529.96	405.34	345.05	426.80
L-242-25	724.5	538.13	736.51	452.48
L-242-38	619.41	687.92	450.99	492.21
L-38	194.96	195.03	285.66	217.95
L-30	340.16	520.01	600.51	400.02
L-35	492.41	442.01	481.07	284.00
L-37	375.00	301.27	259.94	269.94

El Rancho

	Bloques			
Variedades	1	2	3	4
L-18	266.75	666.92	300.71	312.04
L-242-43	408.98	276.25	287.03	951.19
L-242-7	462.64	265.23	193.46	450.74
L-39	277.73	383.88	548.41	320.56
L-246-19	497.13	325.92	621.40	444.88
L-242-22	853.19	154.22	559.49	501.63
L-242-9	549.84	203.22	801.02	392.00
L-246-9	503.10	338.67	147.07	352.18
L-242-25	708.12	797.18	213.28	407.10
L-242-38	182.53	425.06	138.54	440.86
L-38	166.55	151.68	235.65	236.25
L-30	453.70	557.82	372.81	263.11
L-35	523.09	467.08	461.54	407.48
L-37	201.07	361.75	552.94	399.51

Estancia de la Virgen

	Bloques			
Variedades	1	2	3	4
L-18	131.96	90.04	59.62	103.49
L-242-43	185.33	280.97	147.5	240.24
L-242-7	209.25	222.75	112.24	145.57
L-39	165.92	229.05	158.74	124.45
L-246-19	220.12	278.65	232.63	190.80
L-242-22	169.98	160.02	289.52	190.06
L-242-9	426.59	370.1	412.75	399.55
L-246-9	175.23	198.11	152.27	189.24
L-242-25	284.19	199.76	128.65	178.3
L-242-38	155.72	152.88	169.88	141.3
L-38	149.3	189.38	143.22	156.3
L-30	187.67	242.06	193.34	167.51
L-35	170.8	134.32	208.88	235.17
L-37	300.34	336.65	153.94	230.44

6. El siguiente conjunto de datos, se refieren a un experimento en bloques al azar con 8 variedades de papa (*Solanum tuberosum* L.) y 4 bloques. La variable de respuesta medida fue el rendimiento de tubérculos, expresado en $tm.ha^{-1}$.

	Bloques			
Variedades	1	2	3	4
A (Kennebec)	9.2	13.4	11.0	9.2
B (Huinkul)	21.1	27.0	26.4	25.7
C (Sta. Rafaela)	22.6	29.9	24.2	25.1
D (Buena Vista)	15.4	11.9	10.1	12.3
E (B 25-50 E)	12.7	18.0	18.2	17.1
F (B 1-52)	20.0	21.1	20.0	28.0
G (B 116-51)	23.1	24.2	26.4	16.3
H (B 72-53 A)	18.0	24.6	24.0	24.6

a) Construya un conjunto de *box plots* para bloques, analice el comportamiento.
b) Verifique los supuestos del ANOVA.
c) Realice el ANOVA.
d) En caso de ser necesario realice el análisis post-ANOVA.

7. Revise 5 trabajos de graduación o tesis, en lo que se haya utilizado diseño bloques completos al azar. Investigue qué gradiente de variabilidad usaron para definir los bloques y a su criterio indique si estuvo correcto.

8. De un ensayo de introducción de seis especies del género *Eucalyptus*, se tomaron los datos de crecimiento de altura total (en metros) a los diez meses de edad. El ensayo fue establecido en un diseño de bloques completos al azar.

	Bloques			
Tratamientos	1	2	3	4
E. tereticornis	1.95	1.31	1.48	1.60
E. grandis	1.70	1.35	1.39	1.50
E. comphocephala	1.58	1.10	0.91	1.20
E. robusta	1.37	1.01	0.85	1.10
E. citriodora	1.56	1.20	1.08	1.30
E. saligna	1.30	0.90	0.88	1.10

a) Plantee el modelo y defina sus componentes
b) Defina las hipótesis y construya el cuadro de ANOVA. Utilice α=0.05
c) Efectúe la prueba de Tukey para comparar las medias de los tratamientos.

9. Daetz, C.G. (2015), realizó el trabajo de tesis titulado: Evaluación del crecimiento de plantaciones de eucalipto en Lanquín, Alta Verapaz. El experimento se realizó en la finca Setzac, que se encuentra en el municipio de Lanquín, Alta Verapaz. La finca se ubica a 26 kms del casco urbano, a una una altitud de 700 msnm.

Las variables evaluadas en la investigación fueron las siguientes: diámetro a la altura del pecho (DAP en cm), Altura total (m) y volumen por hectárea (m^3).

El diseño experimental utilizado fue el de bloques completos al azar con tres repeticiones. La unidad experimental estuvo formada por 15 plantas (3 surcos × 5 plantas por surco) con un distanciamiento de 3m × 2m, que es equivalente a un área de $90m^2$ por unidad experimental.

A continuación se presentan los resultados a los 12 meses luego de instalado el experimento.

Material	**Bloque**	**Altura**	**DAP**	**Volumen**
CA-30	I	5.09	5.41	6.64
1214	I	4.93	5.76	9.81
1203	I	4.8	5.76	9.77
1084	I	4.36	5.65	4.99
1198	I	5.4	7	5.76
1197	I	4.85	6.35	8.14
1066	I	4.17	4.51	2.43
1846	I	4.93	6.24	11.07
PE-11	I	4.25	4.58	4
Urophylla	I	3.93	4.75	6.38
1188	I	4.12	3.03	3.01
Camaldulensis	I	3.12	3.42	2.95

CA-30	II	4.88	5.97	8.27
1214	II	4.78	5.33	6.16
1203	II	3.63	4.81	4.43
1084	II	4.82	6.75	13.28
1198	II	4.23	6.18	9.5
1197	II	3.94	5.07	3.47
1066	II	4.5	5.51	8.36
1846	II	3.8	4.54	3.36
PE-11	II	3.9	5.24	6.69
Urophylla	II	3.62	4.81	3.87
1188	II	3.14	3.24	1.95
Camaldulensis	II	2.79	3.48	2
CA-30	III	3.71	5.29	5.36
1214	III	3.93	5.25	6.91
1203	III	4.46	6.02	9.28
1084	III	3.54	5.05	5.03
1198	III	3.06	4.19	3.46
1197	III	3.2	4.05	2.66
1066	III	3.27	5.08	4.48
1846	III	3.14	4.26	2.91
PE-11	III	3.42	4.89	4.78
Urophylla	III	3.63	4.63	5.25
1188	III	3.17	4.28	3.68
Camaldulensis	III	3.68	4.76	4.28

Fuente: Daetz, C.G. 2015. Evaluación del crecimiento de plantaciones de eucalipto en Lanquín, Alta Verapaz. Tesis Ing. Forestal. Universidad Rafael Landívar, Campus San Pedro Cláver S.J., Carrera de Ingeniería Forestal con énfasis en Silvicultura y Manejo de Bosques. 37 p. Disponible en: http://recursosbiblio.url.edu.gt/tesiseortiz/2015/06/22/Daetz-Carlos.pdf

a) Plantee el modelo y defina sus componentes
b) Defina las hipótesis y construya el cuadro de ANOVA. Utilice α=0.05
c) Efectúe la prueba de Tukey para comparar las medias de los tratamientos.

5.4 ANÁLISIS POST ANOVA: CONTRASTES

En muchos métodos de comparación múltiple de medias se utiliza la idea de un contraste. Para definir este término, se considerará un experimento con 5 tratamientos, todos con el mismo número de repeticiones (r), en el cual la hipótesis nula Ho:τ_i =0 fue rechazada. Por lo tanto se sabe que los tratamientos influyen en la variable de respuesta, o sea, algunos de los tratamientos producen diferente efecto; sin embargo ¿cuáles son estos tratamientos?

Si se supone que luego de realizar el experimento, los tratamientos 4 y 5 producen el mismo efecto sobre la variable de respuesta, implica que es deseable probar las hipótesis:

$$\text{Ho: } \mu_4 = \mu_5$$
$$\text{Ha: } \mu_4 \neq \mu_5$$

Estas hipótesis pueden ser probadas investigando una combinación lineal apropiada de los totales de tratamientos, por ejemplo:

$$Y_{4.} - Y_{5.} = 0$$

De haber supuesto que el promedio de los tratamientos 1 y 3 no difiere del promedio de los tratamientos 4 y 5, las hipótesis que deben probarse son:

$$\text{Ho: } \mu_1 + \mu_3 = \mu_4 + \mu_5$$
$$\text{Ha: } \mu_1 + \mu_3 \neq \mu_4 + \mu_5,$$

y esto implica la combinación lineal:

$$Y_{1.} + Y_{3.} - Y_{4.} - Y_{5.} = 0$$

En general, la comparación de medias de tratamientos conlleva una combinación lineal de totales de tratamientos de la forma:

$$C = \sum_{i=1}^{t} C_i Y_{i.} ,$$

siendo,

C_i = coeficiente del contraste ortogonal asociado a $Y_{i.}$

$Y_{i.}$ = total del tratamiento i incluido en el contraste;

imponiendo la restricción $\sum_{i=1}^{t} C_i = 0$ a la función lineal C, ésta se denomina: contraste. La suma de cuadrados de un contraste es dada por la siguiente expresión:

$$SC_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.} \right)^2}{r \sum_{i=1}^{t} C_i^2} ,$$

Y tiene un solo grado de libertad.

Para probar un contraste se debe comparar su suma de cuadrados con el cuadrado medio del error experimental. La estadística que resulta tiene una distribución F con 1 y k grados de libertad, siendo k el número de grados de libertad asociado a la suma de cuadrados del error experimental.

5.4.1 Contrastes Ortogonales

Un caso especial del procedimiento anterior es el de los contrastes ortogonales. Dos contrastes con coeficientes C_i y $C_{i'}$ son ortogonales sí: $\sum_{i=1}^{t} C_i C_{i`} = 0$

Si se tiene t tratamientos, el conjunto de t −1 contrastes ortogonales descomponen la suma de cuadrados debida a tratamientos en t–1 componentes independientes con un grado de libertad. Por tanto, las pruebas realizadas sobre los contrastes ortogonales son independientes, lo que significa que no contienen información redundante. En otras palabras, la información que proporciona un contraste no se traslapa con la proporcionada por otro.

¿Cómo obtener los coeficientes para contrastes ortogonales?

Existen muchas maneras de elegir coeficientes de los contrastes ortogonales para un conjunto de tratamientos. Usualmente, algo de la naturaleza del experimento debe sugerir las comparaciones que resultan de interés. Por ejemplo, si se tienen t=3 tratamientos, siendo el tratamiento 1 el control, y los tratamientos 2 y 3 los niveles reales del factor de interés para quien realiza el experimento, los contrastes ortogonales apropiados podrías ser los siguientes:

Tratamientos	Coeficientes para contrastes ortogonales	
1 (control)	–2	0
2 (nivel 1)	1	–1
3 (nivel 2)	1	1

Debe observarse que:

- Contraste 1 con C_i = –2, 1, 1 compara el efecto promedio de los niveles de interés con el control, mientras que:
- Contraste 2 con $C_{i`}$ = 0, –1, 1 compara los dos niveles del factor de interés.

Nota: Los coeficientes de los contrastes deben ser elegidos antes de realizar el experimento y analizar los datos.

5.4.2 Ejemplo de aplicación

Se utilizará el ejemplo de los nematicidas para ilustrar la aplicación de la técnica estadística de contrastes ortogonales, la cual se puede resumir en los siguientes pasos:

1. Definir con anterioridad un conjunto de (t–1) contrastes ortogonales. En este caso, como se tienen 6 tratamientos, los grados de libertad para los tratamientos es de 5, por lo tanto se pueden generar 5 subgrupos, los cuales se describen a seguir:

    1. Testigo (A) contra Nematicidas (B, C, D, E, F)
    2. Carbofuran (E) contra Oxamyl (B,C,D,F)
    3. Oxamyl aplicado al follaje (1.5 y 2.0 lts) contra Oxamyl aplicado al suelo (1.5 y 2.0 lts)
    4. Oxamyl 1.5 lt (aplicado al follaje) contra Oxamyl 2.0 lt (aplicado al follaje)
    5. Oxamyl 1.5 lt (aplicado al suelo) contra Oxamyl 2.0 lt (aplicado al suelo)

2. Definir los coeficientes de los contrastes (C_i)

Contrastes		Tratamientos						$\sum_{i=1}^{t} C_i$
		A	B	C	D	E	F	
C_1	A = B+C+D+E+F	–5	1	1	1	1	1	0
C_2	E = B+C+D+F	0	1	1	1	–4	1	0
C_3	B + D = C + F	0	–1	1	–1	0	1	0
C_4	B = D	0	–1	0	1	0	0	0
C_5	C = F	0	0	–1	0	0	1	0

3. Comprobar la ortogonalidad de los contrastes:

Recuerde que: dos contrastes con coeficientes C_i y $C_{i'}$ son ortogonales si el producto entre los coeficientes de los contrastes, dos a dos, es nulo, o sea, C_i y $C_{i'}$ son ortogonales sí: $\sum_{i=1}^{t} C_i \times C_{i'} = 0$

Si se tienen *t* tratamientos, el conjunto de *t*–1 contrastes ortogonales descomponen la suma de cuadrados debida a los tratamientos en *t*–1 componentes independientes con un solo grado de libertad. Por tanto, las pruebas realizadas sobre los contrastes ortogonales son independientes.

Ejemplo: A continuación se evaluará la segunda condición de ortogonalidad entre los contrastes 1 y 2.

$$\begin{bmatrix} -5 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ -4 \\ 1 \end{bmatrix} = (-5\times 0)+(1\times 1)+(1\times 1)+(1\times 1)+(1\times -4)+(1\times 1)=0$$

Para evaluar la ortogonalidad entre C_1 y C_3, se tiene:

$$\begin{bmatrix} -5 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} = (-5\times 0)+(1\times -1)+(1\times 1)+(1\times -1)+(1\times 0)+(1\times 1)=0$$

y así sucesivamente se va verificando la ortogonalidad por pares de contrastes: C_1 y C_4, C_1 y C_5, C_2 y C_3, C_2 y C_4,. . . , C_4 y C_5.

4. Plantear las subhipótesis. En este caso se trabajarán con los totales de cada tratamiento, identificados como: τ_i, i=1,2,. . ., t

Ho: $\tau_2 + \tau_3 + \tau_4 + \tau_5 + \tau_6 - 5\tau_1 = 0$

Ho: $\tau_2 + \tau_3 + \tau_4 + \tau_6 - 4\tau_5 = 0$

Ho: $\tau_3 + \tau_6 - \tau_2 - \tau_4 = 0$

Ho: $\tau_4 - \tau_2 = 0$

Ho: $\tau_6 - \tau_3 = 0$

5. Obtener el valor numérico de los contrastes, la suma de cuadrados de contrastes (igual a cuadrado medio de contrastes) y evaluar las subhipótesis a través de la prueba de F.

Contrastes	Yi	1	2	3	4	5	6	$r\sum_{i=1}^{t} C_i^2$	$\sum_{i=1}^{t} C_i Y_{i.}$	SC Contrastes	Valor de F
		1756	1238	1507	1121	1145	1003				
C_1	A = B+C+D+E+F	–5	1	1	1	1	1	150	–2766	51005.04	22.32*
C_2	E = B+C+D+F	0	1	1	1	–4	1	100	289	835.21	0.37
C_3	B + D = C + F	0	–1	1	–1	0	1	20	151	1140.05	0.50
C_4	B = D	0	–1	0	1	0	0	10	–117	1368.90	0.60
C_5	C = F	0	0	–1	0	0	1	10	–504	25401.60	11.12*
									SC tratamientos =	**79750.80**	

Cálculos:

$r\sum_{i=1}^{t} C_i^2$

Contraste 1: $5\,(-5^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2) \quad = \quad 5(30) \quad = 150$

Contraste 2: $5\,[0^2 + 1^2 + 1^2 + 1^2 + (-4^2) + 1^2] \quad = \quad 5(20) \quad = 100$

Contraste 3: $5\,[0^2 + (-1^2) + 1^2 + (-1^2) + 0^2 + 1^2] \quad = \quad 5(4) \quad = 20$

Contraste 4: $5\,[0^2 + (-1^2) + 0^2 + 1^2 + 0^2 + 0^2] \quad = \quad 5(2) \quad = 10$

Contraste 5: $5\,[0^2 + 0^2 + (-1^2) + 0^2 + 0^2 + 1^2] \quad = \quad 5(2) \quad = 10$

$\sum_{i=1}^{t} C_i Y_{i.}$

Contraste 1: $[\,(-5)(1756) + (1)(1238) + (1)(1507) + (1)(1121) + (1)(1145) + (1)(1003)] = -2766$

Tarea: obtener los valores de $\sum_{i=1}^{t} C_i Y_{i.}$ para los restantes 4 contrastes.

Suma de cuadrados de contrastes (SC_c): como se tiene 1 grado de libertad por contraste, la suma de cuadrados de contraste es igual a cuadrado medio de contraste.

$$SC_c = CM_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.}\right)^2}{r\sum_{i=1}^{t} C_i^2},$$

Contraste 1:

$$SC_c = CM_c = \frac{(-2766)^2}{150} = 51005.04$$

Valor de F para el contraste 1:

$$F = \frac{CM_c}{CM_{ee}} = \frac{51005.04}{2284.70} = 22.32$$

Valor crítico de F: $F_{c(1,20,0.05)} = 4.35$

6. Emitir las respectivas conclusiones

- Existe efecto diferenciado entre aplicar y no aplicar nematicidas. Se recomienda aplicar nematicidas ya que se reduce la cantidad de nematodos.

- Existen diferencias significativas en el efecto producido por el Oxamyl 1.5 lt y Oxamyl 2.0 lt ambos aplicados al suelo. Se recomienda la aplicación de este último, porque dejó el menor total de nematodos vivos.

5.4.3 Obtención de las sumas de cuadrados de contrastes en el Diseño Completamente al Azar (DCA)

- DCA BALANCEADO

$$SC_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.}\right)^2}{r\sum_{i=1}^{t} C_i^2}, \qquad \begin{cases} \sum_{i=1}^{t} C_i = 0 \\ \sum_{i=1}^{t} C_i C_{i`} = 0 \end{cases}$$

- DCA DESBALANCEADO

$$SC_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.}\right)^2}{\sum_{i=1}^{t} r_i C_i^2}, \qquad \begin{cases} \sum_{i=1}^{t} C_i = 0 \\ \sum_{i=1}^{t} r_i C_i C_{i`} = 0 \end{cases}$$

- DCA CON SUBMUESTREO

a) Cuando se utiliza CMee:

$$SC_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.}\right)^2}{r\sum_{i=1}^{t} C_i^2}, \qquad \begin{cases} \sum_{i=1}^{t} C_i = 0 \\ \sum_{i=1}^{t} C_i C_{i`} = 0 \end{cases}$$

b) Cuando se utiliza CMep:

$$SC_c = \frac{\left(\sum_{i=1}^{t} C_i Y_{i.}\right)^2}{rm\sum_{i=1}^{t} C_i^2}.$$

5.4.4 Ejercicios sobre contrastes ortogonales

1. Los siguientes datos son los resultados de un experimento realizado para determinar si cinco fuentes de nitrógeno difirieron en sus efectos sobre la producción de arroz. Se aplicaron los tratamientos al azar a 20 parcelas en un diseño completamente aleatorizado. La tasa de N era constante y los tratamientos eran: $T_1=Ca(NO_3)_2$, $T_2=Na\ NO_3$, $T_3= NH_4NO_3$, $T_4=(NH_2)_2CO$, $T_5=(NH_4)_2SO_4$.

Tratamientos	Rep. 1	Rep. 2	Rep. 3	Rep. 4
T_1	57.2	51.1	48.5	54.9
T_2	40.6	43.0	52.2	32.3
T_3	36.9	29.0	33.7	37.0
T_4	23.3	23.2	24.4	17.0
T_5	36.8	38.7	31.7	43.6

a) Realice el análisis de varianza, utilizando un nivel de significancia del 5%.

b) Dada la naturaleza de los tratamientos, interesa realizar las siguientes comparaciones:
   i. Tratamientos 1, 2 y 3 versus 4 y 5 (nitratos vs. no-nitratos).
   ii. Tratamientos 1 y 2 versus 3.
   iii. Tratamiento 1 versus 2.
   iv. Tratamiento 4 versus 5. Defina los coeficientes necesarios y realice estos contrastes.

c) ¿Son ortogonales estos contrastes? Si lo son, verifique que la sumatoria de sus sumas de cuadrados es igual a la suma de cuadrados de tratamiento.

2. En un experimento se desea conocer si hay alguna variedad de geranio que produce más flores. Se probaron 5 variedades (B1 y B2, dos híbridos belgas; S1 y S2, dos híbridos canadienses, y Te, una variedad comúnmente usada). Se realizó un DBCA con 4 repeticiones.

Tratamiento	Repetición	No. Flores
B1	1	67
B2	1	50
S1	1	46
S2	1	43
Te	1	36
B1	2	51
B2	2	43
S1	2	29
S2	2	33
Te	2	28

Tratamiento	Repetición	No. Flores
B1	3	48
B2	3	52
S1	3	33
S2	3	34
Te	3	31
B1	4	54
B2	4	32
S1	4	37
S2	4	27
Te	4	33

a) Realice el análisis de varianza.
b) Formule y pruebe los contrastes ortogonales de interés e interprete sus resultados.
c) Concluya con base en los resultados obtenidos.

3. Castillo Fratti, A.J. (1996) realizó el trabajo de tesis titulado: "Evaluación del efecto de la incisión anular sobre la calidad y rendimiento de la fruta de uva de mesa (*Vitis vinifera* L.) en dos localidades del nororiente del país". El material que se utilizó en el ensayo fue plantas adultas de la variedad de uva roja, llamada comúnmente "Uva roja del Jute". Los tratamientos

evaluados fueron: **T_1** = Anillado basal del eje central de la planta, inmediatamente después de la poda; **T_2** = Anillado basal del eje central de la planta, en el momento de la floración (antes de la caída de la caliptra); **T_3** = Anillado en los brazos inmediatamente después de la poda; **T_4** = Anillado en los brazos en el momento de la floración (antes de la caída de la caliptra); **T_5** = Testigo (sin anillado). El diseño experimental de bloques al azar fue utilizado, en cada localidad, com 4 repeticiones. La unidad experimental estuvo constituída por 3 plantas de uva completamente desarrolladas (adultas y podadas). Una de las variables de respuesta medidas fue el peso promedio del racimo individual (PR, expresado en gramos) y el rendimiento ($kg.ha^{-1}$). Los resultados obtenidos en las dos localidades se presentan a continuación:

		Teculután, Zacapa		Chiquimula	
Bloque	Tratamiento	PR	Rendimiento	PR	Rendimiento
1	1	398.2	11502.4	378.1	9661
1	2	398.1	11941.8	383.4	8093.2
1	3	420.2	10737.4	392.3	10024.4
1	4	405.9	11724.8	403.1	9852.6
1	5	403.1	11196.1	398.6	9299.7
2	1	394.2	10510.9	389.3	10812.8
2	2	420.1	11668.3	390.4	8674.7
2	3	415.2	11993.5	389.9	8230.4
2	4	410.2	10481.8	405.9	9019.1
2	5	391.4	11305.9	394.7	9647.3
3	1	383.2	8940.4	380.3	9295.3
3	2	410.3	11851.9	392.2	9150.4
3	3	397.9	10609.6	394.1	9194.7
3	4	391.9	10885	410.6	8667.4
3	5	392.4	9155.1	406.1	9925.9
4	1	391.4	10871.1	401.2	10251.9
4	2	397.9	10167.5	390.2	8670.2
4	3	410.3	10485.4	392.1	9148.1
4	4	402.3	10726.9	391.6	10006.6
4	5	402.5	10732.3	398.7	9302.1

a) Realice el análisis de varianza para cada localidad y variable de respuesta, planteando las hipótesis de interés.

b) Formule y pruebe los siguientes contrastes ortogonales de interés e interprete sus resultados:

b.1) Anillado contra testigo (T_5 vs otros tratamientos)
b.2) Anillado basal contra anillado de brazos (T_1 y T_2 vs T_3 y T_4)
b.3) Anillado basal después de la poda contra Anillado basal en la floración (T_1 vs T_2)
b.4) Anillado de brazo en la poda contra Anillado de brazos en la floración (T_3 vs T_4)

4. Calmo, R. (2012) en busca de sustratos alternativos al Peat-moss evaluó el desarrollo de plántulas de matilisguate (*Tabebuia rose*a Bertol.DC.), en la etapa de vivero, en seis diferentes sustratos, en tubete, en un vivero forestal, utilizando un diseño completamente al azar. Para evaluar el desarrollo de las plántulas se midieron las variables: diámetro de tallo en milímetros, longitud aérea y longitud de raíz en centímetros, de las plántulas de matilisguate. Cada unidad experimental estuvo formada por 24 plantas, colocadas cada una en un tubete de volumen igual a 150 cm^3.

   Los sustratos evaluados fueron los siguientes: Peat-moss (1), tuza de maíz (2), cascarilla de cardamomo (3), aserrín (4), rastrojo de maíz (5) y cascara de coco (6) A continuación se presentan los resultados de la variable longitud de raíz en centímetros de las plántulas de matilisguate.

Tratamiento	Repetición	Longitud
1	1	10.06
1	2	9.78
1	3	9.86
1	4	10.67
1	5	9.73
1	6	9.15
2	1	7.34
2	2	8.68
2	3	6.22
2	4	8.71
2	5	7.57
2	6	7.68
3	1	7.5
3	2	6.68
3	3	8.82
3	4	5.8
3	5	7.54
3	6	5.79

Tratamiento	Repetición	Longitud
4	1	8.23
4	2	8.76
4	3	8.66
4	4	7.42
4	5	9.13
4	6	8.15
5	1	8.85
5	2	7.45
5	3	6.25
5	4	6.9
5	5	9.18
5	6	9.07
6	1	8.41
6	2	7.71
6	3	6.53
6	4	7.43
6	5	8.37
6	6	7.51

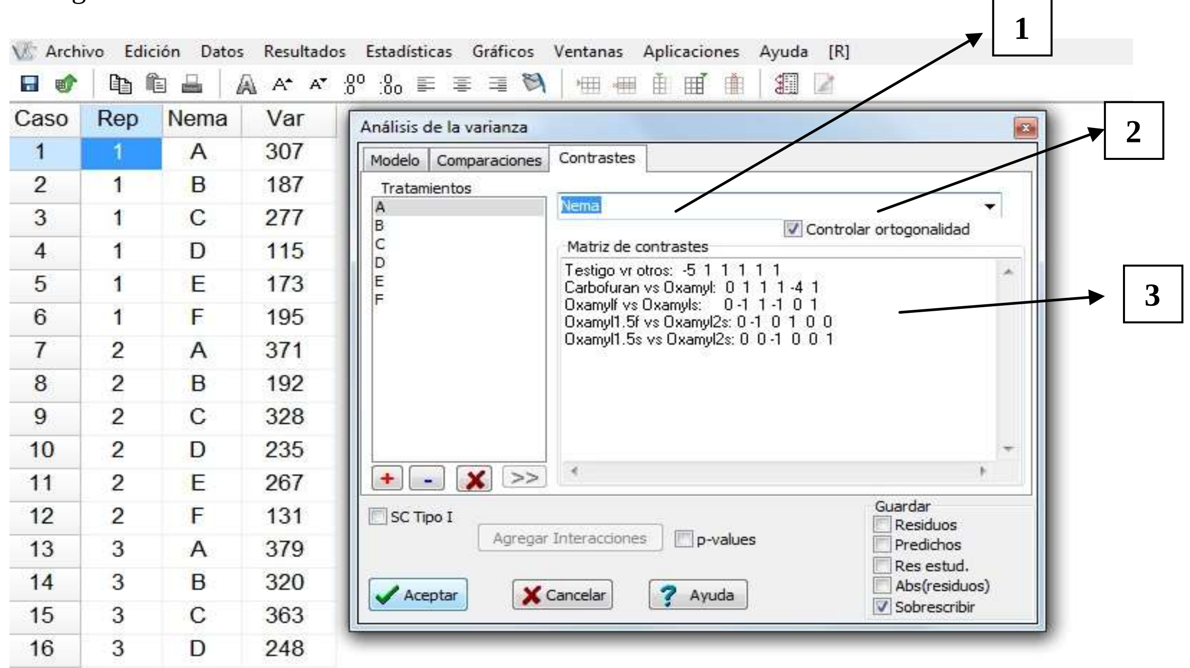
Fuente: Calmo, R. 2012. Evaluación de cinco sustratos alternativos al peat- moss para la produccion de plantulas de matilisguate (*Tabebuia rosea* Bertol. DC.), en tubete, en la comunidad de Salacuim, Cobán, Alta Verapaz, Guatemala. Trabajo de Graduación. USAC, Facultad de Agronomía. Disponible en: http://biblioteca.usac.edu.gt/tesis/01/01_2752.pdf

a) Plantee las hipótesis respectivas
b) Realice el análisis de varianza.
c) Realice el análisis de contrastes ortogonales. Justifique los grupos formados.

5.5 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO BLOQUES COMPLETOS AL AZAR, CON CONTRASTES ORTOGONALES

```
OPTIONS nodate nonumber;
DATA contras;
INPUT rep nema $ var;
CARDS;
1     A     307
1     B     187
1     C     277
1     D     115
1     E     173
1     F     195
2     A     371
2     B     192
2     C     328
2     D     235
2     E     267
2     F     131
3     A     379
3     B     320
3     C     363
3     D     248
3     E     251
3     F     171
4     A     360
4     B     243
4     C     195
4     D     267
4     E     254
4     F     253
5     A     339
5     B     296
5     C     344
5     D     256
5     E     200
5     F     253
;
PROC glm;
    CLASS rep nema;
    MODEL var = rep nema/SS1;
    CONTRAST "T. Abs. vr otros" nema -5 1 1 1 1 1;
    CONTRAST "Carbofuran vs Oxamyl" nema 0 1 1 1 -4 1;
    CONTRAST "Oxamylf vs Oxamyls" nema 0 -1 1 -1 0 1;
    CONTRAST "Oxamyl1.5f vs Oxamyl2f" nema 0 -1 0 1 0 0;
    CONTRAST "Oxamyl1.5s vs Oxamyl2s" nema 0 0 -1 0 0 1;
    LSMEANS nema;
    RUN;
```

Para realizar el procedimiento en Infostat, luego de ingresar los datos y definido el modelo, similar a como fue explicado en el inciso 3.7, seleccione la opción CONTRASTES, como se muestra en el cuadro siguiente:



Algunas observaciones:

1. Seleccione el nombre del factor con el que construirá los contrastes.
2. Verifique que esté activa la opción Controlar ortogonalidad.
3. Una opción es digitar directamente en el cuadro Matriz de contrastes, los nombres de los contrastes y sus respectivos coeficientes.

Luego al dar un clic en Aceptar, obtendremos los resultados siguientes:

Análisis de la varianza

Variable	N	R²	R² Aj	CV
Var	30	0.69	0.56	18.46

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	103228.13	9	11469.79	5.02	0.0013
Rep	23477.33	4	5869.33	2.57	0.0695
Nema	79750.80	5	15950.16	6.98	0.0006
Error	45693.87	20	2284.69		
Total	148922.00	29			

Contrastes

Nema	Contraste	SC	gl	CM	F	p-valor
Testigo vr otros	-553.20	51005.04	1	51005.04	22.32	0.0001
Carbofuran vs Oxamyl	57.80	835.21	1	835.21	0.37	0.5522
Oxamylf vs Oxamyls	30.20	1140.05	1	1140.05	0.50	0.4881
Oxamyl1.5f vs Oxamyl2f	-23.40	1368.90	1	1368.90	0.60	0.4480
Oxamyl1.5s vs Oxamyl2s	-100.80	25401.60	1	25401.60	11.12	0.0033
Total		79750.80	5	15950.16	6.98	0.0006

Compare los resultados con los obtenidos al realizar el ejercicio manualmente y con la salida proporcionada por SAS.

5.6 DISEÑO BLOQUES AL AZAR CON DATOS FALTANTES

Para ejemplificar esta situación, se utilizarán los datos de contenido de proteína (expresados en porcentaje), obtenidos en el experimento sobre amaranto realizado por Alfaro Villatoro (1985).

CASO I: Supongamos que al realizar los análisis de proteína, la muestra correspondiente al corte a 25 días en el 4º bloque fue extraviada.

Cuadro de datos

Bloque	Edad al corte (días)			$Y_{.j}$
	25	40	60	
1	30.6	22.4	13.6	66.6
2	28.5	24.1	13.1	65.7
3	29.1	26.3	15.8	71.2
4	**Y_{14}**	24.3	20.2	**44.5 = B**
5	30.6	18.6	13.8	63.0
6	30.6	20.9	12.0	63.5
7	28.3	21.1	12.7	62.1
8	28.7	23.7	13.7	66.1
$Y_{i.}$	**206.4 = T**	181.4	114.9	**502.7 = S**

a) Ho : $\forall_i$ i= 1, 2, . . . / $\tau_i = \tau$

Ha: $\exists_i$ i =1,2, . . . / $\tau_i \neq \tau$

b) Modelo Estadístico:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3, \ldots, t \\ j = 1, 2, 3, \ldots, r \end{cases}$$

Siendo:

Y_{ij} = contenido de proteína (%) en la i-ésima época de corte y el j-ésimo bloque.
μ = media general del contenido de proteína.
τ_i = efecto de la i-ésima época de corte sobre el contenido de proteína
β_j = efecto del j-ésimo bloque sobre el contenido de proteína.
ε_{ij} = error asociado a la ij-ésima unidad experimental.

c) Supuestos

$\varepsilon_{ij} \sim NID(0, \sigma^2)$
No existe interacción bloque por época de corte

d) Estimación del dato faltante $\hat{Y}_{ij} = \dfrac{tT + rB - S}{(t-1)(r-1)}$

Siendo que:
t = número de tratamientos
T = total del tratamiento donde está el dato faltante.
r = número de repeticiones
B = total del bloque donde está el dato faltante
S = gran total.

Para nuestro caso tenemos que:

$$\hat{Y}_{14} = \frac{3(206.4) + 8(44.5) - 502.7}{(3-1)\,(8-1)} = 33.8$$

e) Análisis de varianza

Los cálculos para el análisis de varianza se hacen como habitualmente, con la única diferencia de que al total y al error experimental se le resta un grado de libertad.

Incorporando la estimativa del dato faltante, tenemos que:

T = 240.2 B = 78.3 y S = 536.5, y las sumas de cuadrados son:

$$\text{SCtot} = (30.6)^2 + (28.5)^2 + \ldots + (13.7)^2 - \frac{(536.5)^2}{24} = 1094.24$$

$$\text{SCbloques} = \frac{(66.6)^2 + (65.7)^2 + \ldots + (66.1)^2}{3} - \frac{(536.5)^2}{24} = 66.74$$

$$\text{SCtrats} = \frac{(240.2)^2 + (181.4)^2 + (114.9)^2}{8} - \frac{(536.5)^2}{24} = 982.49$$

El resumen de análisis de varianza se presenta a continuación:

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	Valor crítico de F
Edad al corte	2	982.49	491.24	141.98*	3.81
Bloques	7	66.74			
Error Experimental	13	45.01	3.46		
Total	22	1094.24			

CV = 8.32 %

Conclusión: El contenido de proteína en las hojas de amaranto varía significativamente con la edad en que la planta se corta.

f) Comparación múltiple de medias de acuerdo con el criterio de Tukey

- Para medias donde no hubo parcelas perdidas, se utiliza el comparador (W)

$$W = q_{(t,glee,\alpha)} \times \sqrt{\frac{CMee}{r}}$$

- Cuando una de las medias a comparar es la que no tuvo pérdida de parcela, W se calcula así:

$$W = q_{(t,glee,\alpha)} \times \sqrt{\frac{1}{2} CMee \left[\frac{2}{r} + \frac{t}{r(r-1)(t-1)} \right]}$$

CASO II: Supongamos ahora que por alguna razón hace falta los datos correspondientes a las observaciones Y_{14} y Y_{35}.

Cuadro de datos

Bloque	Edad al corte (días)			$Y_{.j}$
	25	40	60	
1	30.6	22.4	13.6	66.6
2	28.5	24.1	13.1	65.7
3	29.1	26.3	15.8	71.2
4	**Y_{14}**	24.3	20.2	**44.5**
5	30.6	18.6	**Y_{35}**	**49.2**
6	30.6	20.9	12.0	63.5
7	28.3	21.1	12.7	62.1
8	28.7	23.7	13.7	66.1
$Y_{i.}$	**206.4**	181.4	**101.1**	**488.9**

Estimación de los datos faltantes: Y_{14} y Y_{35}.

a.1) Se debe obtener una estimativa inicial para uno de los datos:

$$\hat{Y}_{14} = \frac{\bar{Y}_{1.} + \bar{Y}_{.4}}{2} = \frac{(206.4/7) + (44.5/2)}{2} = 25.87$$

a.2) Luego se debe suponer que sólo hay un dato faltante y se estima con la siguiente ecuación:

T = 101.1
B = 49.2
S = 488.9 + 25.87 = 514.77

$$\hat{Y}_{35} = \frac{3(101.1) + 8(49.2) - 514.77}{(3-1)\,(8-1)} = 13.01$$

a.3) Se ignora la primera estimación de Y_{14} y se aplica la ecuación:

T = 206.4
B = 44.5
S = 488.9 + 13.01 = 501.91

$$\hat{Y}_{14} = \frac{3(206.4) + 8(44.5) - 501.91}{(3-1)\,(8-1)} = 33.81$$

a.4) Se repite el paso (a.2) usando la nueva estimación de Y_{14}

T = 101.1
B = 49.2
S = 488.9 + 33.81 = 522.71

$$\hat{Y}_{35} = \frac{3(101.1) + 8(49.2) - 522.71}{(3-1)\,(8-1)} = 12.44$$

a.5) Se repite el paso (a.3) con Y_{35} =12.44

T = 206.4
B = 44.5
S = 488.9 + 12.44 = 501.34

$$\hat{Y}_{14} = \frac{3(206.4) + 8(44.5) - 501.34}{(3-1)\,(8-1)} = 33.85$$

Las estimaciones definitivas son: $\hat{Y}_{14} = 33.85$ y $\hat{Y}_{35} = 12.44$

Resumen del análisis de varianza

En el análisis de varianza la única modificación es que disminuye en dos, los grados de libertad para el total y para el error experimental.

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	Valor crítico de F
Edad al corte	2	1005.16	502.58	137.29*	3.89
Bloques	7	71.34			
Error Experimental	12	43.93	3.66		
Total	21	1120.43			

CV = 8.58 %

Conclusión: la misma que en el caso 1.

El programa en SAS para poder obtener las estimaciones de los 2 datos faltantes, se presenta a continuación:

```
DATA w1;
do trat= 1 to 3;
do bloque= 1 to 8;
input prot @;
output;
end;
end;
datalines;
30.6    28.5    29.1    0       30.6    30.6    28.3    28.7
22.4    24.1    26.3    24.3    18.6    20.9    21.1    23.7
13.6    13.1    15.8    20.2    0       12      12.7    13.7
;
/* covariable X, referente a la primera parcela perdida*/
```

```
data w2;
do trat= 1 to 3;
do bloque= 1 to 8;
input X @;
output;
end; end;
datalines;
0	0	0	-1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
;
/* covariable X, referente a la segunda parcela perdida*/

data w3;
do trat= 1 to 3;
do bloque= 1 to 8;
input Y @;
output;
end; end;
datalines;
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	-1	0	0	0
;
/* Concatenación horizontal de los archivos w1, w2 y w3*/

data w; merge w1 w2 w3;
proc print data=w;
title "archivos concatenados";
run;

proc glm data=w;
class trat bloque;
model prot=trat bloque X Y/solution SS1 SS4;
run;
```

****

Los valores de las estimaciones de los valores perdidos, proporcionadas por SAS se presentan a continuación, compare los resultados con los obtenidos con el procedimiento presentado en el texto.

Parameter	**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
X	33.84717949	2.51165199	13.48	<.0001
Y	12.43948718	2.51165199	4.95	0.0003

5.7 ANÁLISIS DE VARIANZA PARA UN DISEÑO BLOQUES AL AZAR CON MUESTREO

Con el objetivo de determinar la cantidad mínima efectiva del atrayente sintético de machos (trimedlure) a fin de mejorar el método de detección de adultos de la mosca del mediterráneo **(***Ceratitis capitata* Wiedemann*)*, se realizó un experimento utilizando trampas tipo Jackson con mechas de algodón como dispensadores.

Los tratamientos consistieron en 0.8, 1.6, 2.6, 3.5 y 7.0 ml de trimedlure aplicados en mechas de tamaño proporcional al volumen. Las trampas se distribuyeron en un área con mosca del mediterráneo mediante un diseño en bloques al azar, colocando dos trampas en cada unidad experimental. En el cuadro 1 se reportan los datos del número total de insectos capturados en cada trampa, 16 días después de su instalación.

Cuadro 1. Número de machos de mosca del mediterráneo capturados bajo diferentes cantidades de atrayente sexual.

Volumen de Trimedlure (ml)	Bloques										$Y_{i..}$
	1	2	3	4	5	6	7	8	9	10	
0.8	55	25	44	63	32	100	83	93	145	110	
	19	13	22	13	15	81	39	23	64	80	
$Y_{1j.}$	**74**	**38**	**66**	**76**	**47**	**181**	**122**	**116**	**209**	**190**	1119
1.6	60	15	35	19	35	122	60	94	110	111	
	31	60	11	25	27	51	29	154	222	162	
$Y_{2j.}$	**91**	**75**	**46**	**44**	**62**	**173**	**89**	**248**	**332**	**273**	1433
2.6	62	39	14	14	39	122	77	138	228	93	
	51	19	40	20	71	66	41	130	97	207	
$Y_{3j.}$	**113**	**58**	**54**	**34**	**110**	**188**	**118**	**268**	**325**	**300**	1568
3.5	82	68	28	27	95	121	34	81	105	273	
	58	24	48	44	11	101	97	46	89	119	
$Y_{4j.}$	**140**	**92**	**76**	**71**	**106**	**222**	**131**	**127**	**194**	**392**	1551
7.0	99	12	72	88	30	117	129	46	80	128	
	104	44	101	222	64	59	11	79	129	104	
$Y_{5j.}$	**203**	**56**	**173**	**310**	**94**	**176**	**140**	**125**	**209**	**232**	1718
$Y_{.j.}$	621	319	415	535	419	940	600	884	1269	1387	7389

(*) Los datos fueron tomados de la tesis de Ing Agr. Salazar Rodríguez, J.A. (1985)

1. Hipótesis
   Ho : $\vee_i \; i = 1,2, \ldots / \tau_i = \tau$
   Ha: $\exists_i \; i = 1,2, \ldots / \tau_i \neq \tau$

2. Modelo Estadístico:

$$Y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ij} + \eta_{ijk} \qquad \begin{cases} i = 1, 2, \ldots, t \\ j = 1, 2, \ldots, r \\ k = 1, 2, \ldots, m \end{cases}$$

Siendo:

Y_{ijk}	=	número de moscas machos de mosca del mediterráneo capturadas en el k-cuadro muestra en la i-ésima época de corte y el j-ésimo bloque.
μ	=	media general del número de moscas machos.
τ_i	=	efecto del i-ésimo volumen de trimedlure
β_j	=	efecto del i-ésimo bloque sobre el número de moscas machos capturadas.
ε_{ij}	=	error experimental asociado a la ij-ésima unidad experimental (error entre).
η_{ijk}	=	error de muestreo asociado a la ij-ésima unidad experimental (error dentro).

3. Supuestos
   $\varepsilon ij \sim NID(0, \sigma^2)$
   $\eta_{ijk} \sim NID(0, \sigma_n^2)$
   No existe interacción bloque por época de corte
4. ANOVA

a) Número de grados de libertad

• Total	=	trm – 1	=	(5)(10)(2) – 1	=	99
• Dosis	=	t – 1	=	5 – 1	=	4
• Bloques	=	r – 1	=	10 – 1	=	9
• Error Experimental	=	(r – 1) (t – 1)	=	(9) (4)	=	36
• Error de Muestreo	=	tr(m– 1)	=	(5×10)(2 – 1)	=	50

b) Suma de cuadrados

$$\text{FC} = \frac{\text{Y...}}{\text{trm}} = \frac{7389^2}{(5)(10)(2)} = 545,973.21$$

$$\text{SC}_{\text{total}} = \sum_{i=1}^{t}\sum_{j=1}^{r}\sum_{k=1}^{m} Y_{ijk}{}^2 - \text{FC}$$

$$= (55^2 + 19^2 + 25^2 + \ldots + 108^2) - 545,973.21 = 278,381.79$$

$$\text{SC}_{\text{trat.}} = \frac{\sum_{i=1}^{t} Y_{i..}{}^2}{rm} - \text{FC}$$

$$= \frac{(1119^2 + 1433^2 + 1568^2 + 1551^2 + 1718^2)}{(10)(2)} - 545,973.21 = 10,096.74$$

$$\text{SC}_{\text{bloques}} = \frac{\sum_{i=1}^{t} Y_{.j.}{}^2}{tm} - \text{FC}$$

$$= \frac{(621^2 + 319^2 + \ldots + 1387^2)}{(10)(2)} - 545,973.21 = 122,086.69$$

$$SC_{ee} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij.}^{\;2}}{m} - FC - SC_{trat} - SC_{bloques}$$

$$= \frac{(74^2 + 38^2 + \ldots + 232^2)}{2} - 545,973.21 - 10,096.74 - 122,086.69 = 62,833.80$$

Resumen del análisis de varianza

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrado medios	Valor de F
Dosis	4	10,096.74	2,524.185	$F_2 = 1.485^{NS}$
Bloques	9	122,086.69		
Error Experimental	36	62,833.80	1,745.39	$F_1 = 1.0467^{NS}$
Error de Muestreo	50	83,374.50	1,667.496	
Total	99	278,391.79		

* significativo al 5% de significancia.

$$F_1 = \frac{CM_{ee}}{CMem} = \frac{1745.39}{1667.496} = 1.0467$$ Valor crítico de $F_{1(36,50,0.05)} = 1.65$

1º Pruebas preliminares de significancia

Como $F_1 < F$ (glee, glem, α), se acepta Ho: $\sigma_e^2 = 0$, lo cual indica que el muestreo no fue efectivo o no es importante en este experimento. Por lo que los errores deben mancomunarse de la siguiente manera:

$$CMep = \frac{SCee + SCem}{glee + glem} = \frac{62833.8 + 83374.5}{36 + 50} = 1700.097$$

2º. Prueba definitiva.

$$F_2 = \frac{CM_{trat}}{CM_{ep}} = \frac{2524.85}{1700.097} = 1.485$$ Valor crítico de $F_{2(4,86,0.05)} = 2.40$

$$CV = \frac{\sqrt{CMep}}{\bar{Y}_{\ldots}} \times 100 = \frac{\sqrt{1700.097}}{73.89} \times 100 = 55.80\%$$

¿Cómo obtener las variancias estimadas?
Una manera es utilizando el método de los momentos (o ANOVA)

$$\sigma_e^2 = \frac{CM_{ee} - CM_{em}}{k} = \frac{1745.39 - 1667.496}{2} = 38.95 \text{ e } \sigma_d^2 = 1667.496$$

5.8 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO BLOQUES COMPLETOS AL AZAR, CON MUESTREO

```
OPTIONS nodate nonumber;
DATA dbam;
INPUT dos rep nm;
CARDS;
0.8     1     55
0.8     1     19
1.6     1     60
1.6     1     31
2.6     1     62
2.6     1     51
3.5     1     82
3.5     1     58
7       1     99
7       1     104
..      ..    ..
..      ..    ..
..      ..    ..
..      ..    ..
0.8     10    110
0.8     10    80
1.6     10    111
1.6     10    162
2.6     10    93
2.6     10    207
3.5     10    273
3.5     10    119
7       10    128
7       10    104
;
PROC glm;
CLASS dos rep; /*con muestreo*/
MODEL nm =rep dos rep*dos/SS1;
TEST h=dos E=rep*dos; /*rep*dos representa el error experimental*/
RUN;
PROC glm;
TITLE "mancomunando errores";/*sin efecto de muestreo*/
CLASS dos rep;
MODEL nm=rep dos/ss1;
RUN;
```

Otra alternativa, es digitando luego del conjunto de datos, esta parte del programa:

```
PROC GLM DATA=dbam;
CLASS dos rep;
MODEL nm=dos rep dos*rep/SS1;
RANDOM dos*rep/TEST;/*se define el error experimental como la interacción (dos*rep)*/
RUN;
```

O digitar:

```
/*Análisis de Varianza para un Modelo Mixto*/
PROC MIXED DATA=dbam;
CLASS dos rep;
MODEL nm = dos rep; /*se colocan los efectos fijos involucrados en el modelo*/
RANDOM dos*rep; /*por defecto el error de submuestreo es el que aparece en la salida como error*/
RUN;
```

A continuación se presenta el procedimiento en Infostat.

1. Ingrese los datos a Infostat, de la misma manera como se ingresaron en SAS.

2. En el menú ESTADÍSTICAS ---→ ANÁLISIS DE LA VARIANZA, digitar en variables dependientes: nm y en variables independientes: dos y rep.

3. Para determinar si el muestreo fue efectivo o no, especifique los siguientes términos del modelo: dos, rep y dos*rep. Infostat generará el siguiente cuadro de ANOVA:

Análisis de la varianza

Variable	N	R^2	R^2 Aj	CV
nm	100	0.70	0.41	55.26

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.		SC	gl	CM	F	p-valor
dos		10096.74	4	2524.18	1.51	0.2124
rep		122086.69	9	13565.19	8.14	<0.0001
dos*rep	**(ee)**	62833.86	36	1745.39	1.05	0.4349 NS
Error	**(em)**	83374.50	50	1667.49		
Total		278391.79	99			

4. Como el efecto del submuestreo no fue significativo, los errores deben mancomunarse. Para ello, debe ejecutar de nuevo el programa en Infostat. Retire dos*rep en la especificación de los términos del modelo. Infostat generará el siguiente cuadro de ANOVA:

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
dos	10096.74	4	2524.19	1.48	0.2139
rep	122086.69	9	13565.19	7.98	<0.0001
Error	146208.36	86	1700.10	**error ponderado (o mancomunado)**	
Total	278391.79	99			

5. En aquellos casos en que el error de muestreo es significativo, en la especificación del modelo digite: dos\dos*rep, dos*rep, rep. Infostat realizará el cálculo de la estadística F de tratamientos, considerando el CM del error experimental (1745.39). El resultado será:

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor	(Error)
dos	10096.74	4	2524.18	1.45	0.2389	(dos*rep)
dos*rep	62833.86	36	1745.39	1.05	**error experimental**	
rep	122086.69	9	13565.19			
Error	83374.50	50	1667.49			
Total	278391.79	99				

5.9 EJERCICIOS PROPUESTOS

1. Se desea determinar el efecto de cuatro dietas sobre la ganancia de peso (kg.) postdestete de cabritos, durante una prueba de 200 días. Se utilizaron 32 cabritos en 16 corraletas (dos cabritos por corraleta) distribuyéndose los tratamientos por bloques. Se midió la ganancia en peso durante la prueba para cada cabrito. Los resultados se presentan a continuación:

		Bloque			
Dieta	Cabrito	I	II	III	IV
A	1	63.0	62.8	63.5	740.3
	2	62.7	62.0	64.0	72.5
B	1	51.2	60.9	54.7	54.3
	2	52.0	59.9	69.5	55.6
C	1	47.9	56.7	56.2	54.3
	2	44.8	52.8	51.2	54.8
D	1	38.9	32.9	32.2	34.9
	2	37.2	41.5	35.9	33.8

Con esta información, efectúe el análisis de varianza y en caso de ser necesario, el análisis post-ANOVA.

2. En un experimento se evaluaron cinco atrayentes alimenticios en trampa Mcphail, para detección de mosca de la fruta en mango *(Manguifera indica* L.). Los tratamientos consistieron en:

A. Vinagre de mango + melaza + agua
B. Vinagre de piña + melaza + agua
C. Melaza + agua
D. Proteína hidrolizada + agua
E. Incaparina + agua.

Las trampas se distribuyeron mediante un diseño en bloques al azar, colocando dos trampas en cada unidad experimental. El cuadro siguiente se reportan los datos del número total de insectos capturados por trampa.

Trats.	Repeticiones							
	1	2	3	4	5	6	7	8
A	95	35	60	90	30	100	120	50
	105	55	101	200	70	60	20	80
B	55	25	44	60	30	80	83	90
	20	15	20	15	10	75	40	25
C	55	15	40	20	30	100	65	90
	25	60	10	25	25	50	25	130
D	75	45	25	15	40	120	75	135
	55	20	55	35	80	70	45	120
E	80	15	70	80	30	120	100	40
	50	45	105	200	65	60	15	75

Para los resultados del experimento anterior realice el ANÁLISIS DE VARIANZA.

3. Navarro (2003) evaluó el efecto de cuatro láminas de riego sobre el rendimiento de Plátano (*Musa paradisíaca* var. Curraré) bajo las condiciones de la Aldea "Los Encuentros", Coatepeque, Quetzaltenango. Como las unidades experimentales fueron demasiado grandes, se llevó a cabo un submuestreo. Se tomaron cinco submuestras (1 submuestra = 1 planta de plátano). Una de las variables de respuesta medida fue el rendimiento del fruto de plátano (en kg/ha). Los resultados por tratamiento y bloque se presentan a continuación:

Lámina total aplicada, en mm	Muestra	Bloque			
		I	II	III	IV
345.63	1	21950.11	34421.76	24943.31	36417.23
	2	47891.15	26938.77	19954.64	32426.30
	3	27437.64	31927.43	25941.04	32426.30
	4	21451.24	29931.97	37913.83	37913.83
	5	29931.97	33922.90	29931.97	33424.03
626.08	1	30929.70	28934.24	34920.63	41904.76
	2	39909.29	38911.56	35918.36	36916.09
	3	34920.63	45895.69	27936.50	39909.29
	4	31927.43	39916.09	43900.22	40907.02
	5	29931.97	39909.29	27936.50	44897.95
830.28	1	38911.56	36417.23	40408.16	43900.22
	2	41904.76	37414.96	44399.00	40408.16
	3	44399.09	45895.69	46394.55	45895.69
	4	42403.62	31428.57	34920.63	38412.69
	5	54875.28	37015.80	28435.37	40907.02
1,003.20	1	45895.69	41904.76	50884.35	45396.82
	2	36018.14	42902.49	51882.08	35918.36
	3	29433.10	31927.43	39410.43	44399.09
	4	39909.209	30929.70	33922.90	39909.29
	5	35918.36	44897.95	35419.50	38911.56

Con esta información, efectúe el análisis de varianza y en caso de ser necesario, el análisis post-ANOVA.

4. Un Ingeniero Agrónomo realizó un experimento en Chichicastenango, que consistió en asperjar hojas de manzana (*Pyrus malus*) con tres concentraciones de un compuesto orgánico nitrogenado, luego determinó la cantidad de nitrógeno (mg/dm^2) que permaneció en las hojas inmediatamente después de la aplicación y al final de ciertos tiempos pre-establecidos. La finalidad de este experimento fue determinar la rapidez con la que el nitrógeno es absorbido por las hojas, se realizaron dos reproducciones de cada tratamiento como se muestra en la tabla siguiente:

Tiempo	Concentraciones de nitrógeno		
	N_1	N_2	N_3
T_o	2.29	6.80	8.75
	2.54	5.94	9.52
T_1	0.46	3.03	2.49
	0.19	1.00	2.04
T_2	0.05	0.75	1.40
	0.26	1.16	1.81

Considerando un bloqueo por tiempos, realice el análisis de varianza y emita las conclusiones en términos estadísticos y prácticos.

5. En un ensayo realizado por el Dr. Joaquim Teófilo Sobrinho en la Estación Experimental de Limeira (São Paulo), del Instituto Agronómico de Campinas (IAC), fueron evaluados 13 clones de naranja Pêra-do-Rio. Los resultados de producción, en kg de frutos por planta, en el año 1987 (plantas con 16 años de edad) son presentados a continuación:

Clones	Bloque 1		Bloque 2		Bloque 3	
	Planta 1	Planta 2	Planta 1	Planta 2	Planta 1	Planta 2
Umbigo	36.50	32.40	28.10	35.60	39.80	45.10
Pé Franco	71.40	109.70	62.80	57.00	102.60	23.10
Premunizada	104.90	72.00	59.20	89.20	38.70	103.10
Ipiguá CV2	71.20	58.00	91.60	93.60	44.30	26.20
Messias CV	73.20	47.20	47.80	50.00	50.40	29.90
Santa Irene CN	87.70	41.30	45.70	73.00	49.20	70.20
Tardia CV4	74.20	5.60	18.20	20.10	23.40	28.30
Ipiguá CV1	41.60	57.30	41.50	26.00	32.50	27.20
Bianchi	85.20	66.30	79.00	82.10	51.20	98.60
Santa Tereza	36.90	30.10	29.60	28.10	31.30	31.80
Paulo Rosa	57.30	23.00	29.50	0.00	14.50	34.20
Tardia CV3	10.00	12.10	4.40	15.50	4.00	9.50
Cassiano 2	59.70	33.80	2.40	11.30	10.20	14.40

Fuente: Barbin, D. (2013)

a) Realice el análisis de varianza, incluyendo la fuente de variación: "entre plantas dentro de parcelas".

b) Realice el análisis de varianza con las producciones medias de las parcelas.

c) Aplique la prueba de Tukey en la comparación de medias de los tratamientos (para los incisos anteriores)

6. Un estudiante de la Escuela Nacional Central de Agricultura (ENCA), contó el número de brotes en hule por piso foliar (puede considerarlos como bloques), con dos métodos. En cada piso realizó 4 lecturas (submuestras). La variable respuesta medida fue el número promedio de brotes. Evalue si existen diferencias significativas entre los 2 métodos utilizados.

Método	Piso	brotes	Método	Piso	Brotes
1	1	1.2500	2	1	3.6500
1	1	1.5000	2	1	4.4000
1	1	1.7500	2	1	4.5500
1	1	2.0000	2	1	4.5500
1	2	0.1930	2	2	0.2308
1	2	0.2632	2	2	0.2308
1	2	0.2632	2	2	0.2564
1	2	0.3158	2	2	0.2564
1	3	0.0606	2	3	0.1724
1	3	0.0909	2	3	0.1724
1	3	0.0909	2	3	0.1724
1	3	0.0909	2	3	0.1724

7. En un lote experimental ubicado en Casas Blancas, Michoacán se evaluaron 7 tratamientos de fertilización empleando un diseño de bloques al azar, con la finalidad de estudiar el efecto de la gallinaza en combinación con otros elementos, sobre la composición química del suelo. Las variables de respuesta registradas fueron las cantidades de calcio, magnesio, sodio, potasio y nitrato en partes por millón (ppm) que contenía el suelo después de un período de haber aplicado los tratamientos; en este ejercicio, se utiliza sólo el contenido de nitrato (NO_3) en ppm como variable respuesta. Los tratamientos se repitieron en 3 bloques y se tomaron 3 submuestras por unidad experimental. Los tratamiento aplicados se presentan en la tabla siguiente:

Tratamiento	Nitrógeno (kg)	Fósforo (kg)	Gallinaza (ton)
1	150	0	0
2	150	400	0
3	150	400	0
4	150	400	5
5	150	400	20
6	0	0	20
7	0	0	0

Con los resultados siguientes, realice el ANOVA y concluya.

		Bloque		
Trats	Obs	I	II	III
1	1	76	158	174
1	2	80	158	176
1	3	76	156	179
2	1	66	174	162
2	2	76	175	168
2	3	66	176	162
3	1	76	152	178
3	2	98	146	174
3	3	98	138	176
4	1	147	148	160
4	2	156	148	160
4	3	148	148	150
5	1	140	175	194
5	2	140	174	188
5	3	141	173	188
6	1	140	195	178
6	2	164	195	178
6	3	140	195	178
7	1	100	102	156
7	2	84	108	156
7	3	87	108	156

Fuente: Zamudio, F.; Alvarado, A. 1996. **Análisis de Diseños Experimentales con igual número de submuestras**. Universidad Autónoma Chapingo, División de Ciencias Forestales. 58 p. Disponible en: http://www.geocities.ws/a_alvaseg/submuestreov1.pdf; consultado el 20/12/2013.

8. A continuación se presentan los resultados de un experimento en pastos. Los valores se refieren a los pesos expresados en kilogramos de masa de materia seca cosechadas sobre la misma variedad en partes iguales de parcelas. Con excepción del tratamiento 4 que fue el control, los otros estuvieron representados por diferentes abonos foliares:

	Tratamientos									
Bloque	1		2		3		4		5	
I	7.5	4.5	12.5	13.2	7.0	1.0	1.5	2.0	28.0	29.0
II	15.5	14.0	20.0	18.5	10.0	8.0	13.0	15.0	19.5	16.0
III	16.5	14.5	15.0	14.0	15.5	14.0	8.5	9.0	10.5	12.0
IV	19.0	18.6	23.8	24.4	17.8	18.5	14.8	16.6	22.0	24.8

Con esta información, efectúe el análisis de varianza y en caso de ser necesario, el análisis post-ANOVA. Recuerde plantear el modelo estadístico matemático asociado a este diseño.

9. Considere un experimento en el que se evaluaron 4 variedades de frijol en un diseño de bloques completos al azar, con 5 repeticiones y 2 submuestras por parcela. La variable de respuesta fue el rendimiento expresado en kilogramos de grano por unidad de 10 metros cuadrados.

	Bloques				
Variedades	**I**	**II**	**III**	**IV**	**V**
A	32.3	41.3	29	29.3	22.5
	34.5	42.2	29.5	29.5	22.3
B	33.2	39	28.7	31.4	20.1
	33.8	38.4	27.7	30.2	19.8
C	30.5	35.5	27.8	25.5	17.6
	30.4	34.6	26.5	24.2	15.4
D	29.3	32	25.6	21	11.3
	29	31.5	25.7	22.3	11

a) Describa el modelo estadístico matemático asociado a este experimento.
b) Plantee las hipótesis estadísticas a evaluar
c) Realice el análisis de varianza
d) En caso de ser necesario, aplique la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey.

Fuente: Quiroga, V. 1976. Manual práctico para el análisis de experimentos de campo**.** Instituto Interamericano de Cooperación para la Agricultura**,** Programa de Información Agropecuaria del Istmo Centroamericano. 113 p. Disponible en: https://books.google.com.bo/books?id=i3AOAQAAIAAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Consultado el 27/04/2016.

5.10 EFICIENCIA DEL DISEÑO EN BLOQUES COMPLETOS AL AZAR

5.10.1 Introducción

Este procedimiento es utilizado para verificar la eficiencia de este diseño con relación al diseño completamente al azar, indicando si realmente era necesario la construcción de bloques, y está expresado por:

$$E = \frac{\hat{V}_1(\overline{y}_{i.} - \overline{y}_{i'.})}{\hat{V}_2(\overline{y}_{i.} - \overline{y}_{i'.})},$$

siendo $V_1\left(\overline{y}_{i.} - \overline{y}_{i'.}\right)$ correspondiente a la estimación de la varianza de la diferencia de dos medias referentes al diseño completamente al azar, siendo:

$$\hat{V}_1(\overline{y}_{i.} - \overline{y}_{i'.}) = \frac{2\,CM\,Residuo_1}{r},$$

donde r es el número de repeticiones, y

$$CM\,Residuo_1 = \frac{(SCTotal - SCTrat)}{t(r-1)} = \frac{SC\,Residuo_1}{t(r-1)}$$

$V_2\left(\overline{y}_{i.} - \overline{y}_{i'.}\right)$ corresponde a la estimación de la varianza de la varianza de la diferencia de dos medias referentes al diseño de bloques completos al azar, siendo:

$$\hat{V}_2(\overline{y}_{i.} - \overline{y}_{i'.}) = \frac{2\,CM\,Residuo_2}{r},$$

siendo r el número de repeticiones, y

$$CM\,Residuo_2 = \frac{(SCTotal - SCTrat - SCBloques)}{(t-1)\,(r-1)} = \frac{SC\,Residuo_2}{(t-1)\,(r-1)}.$$

Así, se obtiene el cociente:

$$E = \frac{CM\,Residuo_1}{CM\,Residuo_2},$$

que es denominado de eficiencia relativa para el diseño en bloques completos al azar con relación al diseño completamente al azar, se expresa en porcentaje, indicando cuanto de eficiencia el diseño en bloques completos al azar aumentó sobre el diseño completamente al azar.

5.10.2 Ejemplo de aplicación

Los datos que se presentan a continuación se refieren al rendimiento expresado en toneladas de caña por hectárea (TCH) obtenido en un experimento realizado por González, B.H. (1999) para evaluar alternativas de intercalamiento, en la finca Bougambilia del Ingenio Magdalena, en el municipio de La Democracia, Escuintla. El diseño utilizado fue el de bloques completos al azar.

Modalidades de intercalamiento	Bloques				
	1	2	3	4	5
Caña intercalada con girasol Z-1296 a 15 cm	87.62	106.20	78.81	72.38	85.71
Caña en monocultivo	115.00	126.19	105.12	110.48	97.81
Caña intercalada con girasol Z-1296 a 25 cm	105.48	121.90	130.95	110.00	83.81
Caña intercalada con girasol J-1296 a 15 cm	111.91	105.71	98.10	83.33	141.43
Caña intercalada con girasol J-1296 a 25 cm	105.71	100.95	114.76	123.80	96.67

Con estos datos se realizó un ANOVA, dando como resultado:

F.V.	G.L.	S.C.	C.M.	Valor de F	F crítica	Valor p
Alternativas de interca-Lamiento	4	2204.20	551.05	2.07	3.01	0.132
Bloques	4	460.71	115.18	0.43	3.01	0.782
Residuo (2)	16	4251.21	265.70			
Total	24	6916.12				

Confrontando el valor obtenido para F con el valor de F crítico, no existe evidencia suficiente para rechazar la hipótesis nula a un nivel de α=0.05 de significancia, concluyendo que las diferentes alternativas de intercalamiento no afectan la producción de caña por hectárea, considerando las condiciones particulares edáficas y climáticas del sitio experimental.

Cálculo de la Eficiencia:

$$E = \frac{\text{CM Residuo}_1}{\text{CM Residuo}_2} = \frac{\text{Cuadrado medio del error experimental estimado para un diseño Completamente al Azar}}{\text{Cuadrado medio del error experimental estimado para un diseño Bloques Completos al Azar}}$$

$$\text{CM Residuo}_1 = \frac{(6916.12 - 2204.20)}{5(5-1)} = \frac{4711.922}{20} = 235.5961$$

$\text{CM Residuo}_2 = 265.70$, de esta manera, la eficiencia es:

E = (235.5961/265.70) = 0.89 u 89%

Se aprecia que no se obtuvo ganancia apreciable en la eficiencia por la formación de bloques y el uso del análisis en bloques completos al azar (BCA), ya que el valor de E fue inferior al 100%. Esto es, aparte del hecho de "seguridad" del diseño BCA, el esfuerzo adicional (de hacer bloques) no fue fructífero (Ostle, 1992).

Observación:

Si por ejemplo, el valor de E hubiese sido de 102.05 %, se concluiría que el diseño en bloques completos al azar incrementó 2.05% de eficiencia relativa al diseño completamente al azar.

CAPÍTULO 6
USO DE LA REGRESIÓN EN EL ANÁLISIS DE VARIANZA: POLINOMIOS ORTOGONALES

6.1 INTRODUCCIÓN

Cuando se planean experimentos, puede darse el caso que los niveles de los factores o los tratamientos sean cuantitativos. Por ejemplo, cuando se evalúa la aplicación de dosis crecientes de un fertilizante en un determinado cultivo con el objetivo de observar su efecto sobre la producción. Si además, los niveles o tratamientos de interés tienen incrementos igualmente espaciados, es interesante conocer la naturaleza de la respuesta más que la comparación entre tratamientos o niveles; para ello se recurre a la técnica de polinomios ortogonales.

Un polinomio ortogonal es una ecuación de regresión. Así, se puede suponer que existe una variable dependiente o variable de respuesta, representada por y, dependiendo de k variables independientes o predictoras, por ejemplo: $X_1, X_2, \ldots, X_k$. Esa relación entre las variables es caracterizada por un modelo matemático–estadístico. Este modelo es fijo para un conjunto de datos. Si el investigador conoce la forma verdadera de la relación funcional existente entre y y $X_1, X_2, \ldots, X_k$, puede considerar que:

$$y = f(X_1, X_2, \ldots, X_k),$$

en la mayoría de los casos, esa relación funcional es desconocida, y el investigador puede seleccionar una función apropiada para aproximar f. Los modelos polinomiales son ampliamente utilizados como funciones aproximadas de f, y en ellos está incluido el modelo de regresión lineal simple.

6.2 MODELOS DE REGRESIÓN POLINOMIAL

El modelo de regresión polinomial de grado p para una única variable independiente es representado por:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_p X_i^p + \varepsilon_i \qquad (1)$$

con:

$i = 1, 2, \ldots, n$

$\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_p$ son los parámetros de la regresión que serán estimados.

$\varepsilon_i \sim N(0, \sigma^2)$ y son independientes.

Suponiendo n pares de datos $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ y todos los niveles referentes a la variable X son equidistantes, es decir: $X_1 = X_1$; $X_2 = X_1+q$; $X_3 = X_2+q$; $\ldots$, $X_n = X_{n-1}+q$. El modelo (1) entonces, puede ser escrito de la siguiente manera:

$$y_i = b_0 + b_1 P_1(X_i) + b_2 P_2(X_i) + \ldots + b_p P_p(X_i) + \varepsilon_i ,$$

siendo:

$i = 1, 2, \ldots, n$

$\varepsilon_i \sim N(0, \sigma^2)$ y son independientes.

$b_0, b_1, b_2, b_3, \ldots, b_p$ son las estimaciones de los parámetros de la regresión.

$P_k(X_i)$ es un polinomio ortogonal de orden $k = 1, \ldots, p$; siendo:

(i) $P_0(X_i) = 1$;

(ii) $\sum_{i=1}^{n} P_k(X_i) = 0$;

(iii) $\sum_{i=1}^{n} P_k(X_i) \times P_{k'}(X_i) = 0$, para $k \neq k'$;

(iv) $\sum_{k=1}^{p} P_k^{\,2}(X_i) \neq 0$

Los valores de $P_k(X_i)$, $k = 1, \ldots, p$; cuando los niveles de la variable X son equidistantes, pueden ser obtenidos a través de las siguientes expresiones, encontradas en Gomes (2000):

$$P_1(X_i) = x_i\ ; \quad \text{siendo} \quad x_i = \frac{X_i - \overline{X}}{q}\ ;$$

siendo que:

X_i son los valores de la variable predictora o independiente,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\ ;$$

q = amplitud entre dos niveles consecutivos.

$P_2(X_i) = x_i^{\,2} - \frac{n^2 - 1}{12}$, siendo n = número de variables independientes.

$$P_3(X_i) = x_i^{\,3} - \frac{3n^2 - 7}{20} \times x_i\ ,$$

$$P_4(X_i) = x_i^{\,4} - \frac{3n^2 - 13}{14} \times x_i^{\,2} + \frac{3(n^2 - 1)\,(n^2 - 9)}{560}\ ,$$

$$P_5(X_i) = x_i^{\,5} - \frac{5(n^2 - 7)}{18} \times x_i^{\,3} + \frac{15n^4 - 230n^2 + 407}{1008} \times x_i\ .$$

A través de la aplicación del método de los mínimos cuadrados, se obtiene $\hat{b}_k$, un estimador de mínimos cuadrados de b_k, con $k = 1, \ldots, p$; cuya ecuación es:

$$b_k = \frac{\sum_{i=1}^{n} P_k(X_i)\ y_{i.}}{r\sum_{i=1}^{n} P_k^{\,2}(X_i)}\ ,$$

siendo:

$P_k(X_i)$ = coeficientes del polinomio ortogonal de grado k asociado al nivel del factor.

$y_{i.}$ = total del nivel i del factor en estudio.

r = número de repeticiones.

Las hipótesis a ser evaluadas son:

H_0: $b_k = 0$ contra H_a: $b_k \neq 0$, a través de la estadística:

$$F_o = \frac{CM\ Reg_k}{CM\ Resíduo}\ ,\ \text{siendo:}$$

$$CM\ Reg_k = \frac{SC\ Reg_k}{1},\ \text{y}\ \ SC\ Reg_k = \frac{\left(\sum_{i=1}^{n} P_k(X_i)\ y_{i.}\right)^2}{r\sum_{i=1}^{n} P_k^{\,2}(X_i)},\ \text{asociada a 1 grado de libertad,}$$

y se rechaza H_0 con un nivel α de significancia cuando $F_o \geq F_{(1,m,\,\alpha)}$.

Con estas consideraciones y aplicando el teorema de Cochran (Neter, 1996, páginas 76-77) se tiene que:

(i) $\frac{SC\ Reg_k}{\sigma^2}$, bajo la hipótesis nula, tiene distribución χ^2 con 1 grado de libertad.

(ii) $\frac{SC\ Residuo}{\sigma^2}$, tiene distribución χ^2 con m grados de libertad.

(iii) SC Reg_k y SC Residuo son independientes.

Entonces, el cociente entre esas variables aleatorias, o sea,

$$\frac{SC\ Reg_k / 1}{SC\ Residuo / m} = \frac{CM\ Reg_k}{CM\ Residuo} = F_o$$

que tiene distribución de F (Fisher–Snedecor) con 1 y m grados de libertad.

6.3 EJEMPLO DE APLICACIÓN

En un experimento de fertilización en Eucalipto (*Eucalyptus camaldulensis*), realizado en vivero, fueron usadas 4 dosis de potasio (0, 30, 60 y 90 ppm), obteniéndose las alturas (en cm) presentadas en el cuadro 1.

El diseño utilizado fue: completamente al azar, con 3 repeticiones. Se solicita:

a) Obtener los polinomios ortogonales hasta el 3er. grado.

b) Verificar el ajuste de un polinomio de 3er grado a esos datos.

c) Obtener la ecuación de regresión polinomial más adecuada.

Cuadro 1. Alturas (expresadas en cm) de plantas de Eucalipto.

Dosis K	Repeticiones			$y_{i.}$	$\bar{y}_{i.}$
	I	II	III		
0	80	86	71	237	79.00
30	144	151	97	392	130.67
60	151	127	117	395	131.67
90	70	85	92	247	82.33

Solución:

El modelo adoptado para el análisis de varianza fue el siguiente:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \ ,$$

con i = 1,2, 3, 4 y j = 1, 2, 3.; siendo:

Y_{ij} el valor observado referente a la i-ésima dosis de potasio aplicada en la j-ésima repetición.
τ_i es el efecto de la i-ésima dosis de potasio,
ε_{ij} es el error experimental asociado a Y_{ij}, tal que $\varepsilon_{ij} \sim N(0, \sigma^2)$ e independiente.

A continuación se presentan as hipótesis a ser evaluadas:

H_0: $\tau_i = 0$, para todo i, contra
H_a: $\tau_i \neq 0$, para algún i,

por medio de la prueba de F, aplicado al análisis de varianza, cuyos resultados se encuentran en el cuadro 2.

Cuadro 2. Análisis de Varianza con prueba F

F.V.	**GL**	**SC**	**CM**	**Valor de F**	**F crítica**
Dosis	3	7668.92	2556.31	7.57 *	4.07
Residuo	8	2702.00	337.75		
Total	11	10370.92			

Se rechaza H_0 con un nivel de 5% de significancia, concluyendo que existe efecto de dosis crecientes de potasio sobre la altura de las plantas de Eucalipto. Continuando con el análisis estadístico, se efectuó el análisis de regresión polinomial, considerando el modelo de regresión polinomial de tercer grado, representado por el modelo que se muestra a continuación:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i \text{ , o}$$

$$y_i = b_0 + b_1 P_1(X_i) + b_2 P_2(X_i) + b_3 P_3(X_i) + \varepsilon_i$$

con i = 1, 2, 3, 4.

Las hipótesis a ser evaluadas:

H_{01}: $b_1 = 0$ contra H_{a1}: $b_1 \neq 0$,
H_{02}: $b_2 = 0$ contra H_{a2}: $b_2 \neq 0$,
H_{01}: $b_3 = 0$ contra H_{a3}: $b_3 \neq 0$,

Usando la prueba F y aplicando el análisis de varianza, llegamos a los resultados que se muestran en el cuadro 3.

Cuadro 3. Análisis de regresión polinomial con aplicación de la prueba F

F.V.	GL	SC	CM	Valor de F	F crítica
Reg. Lineal	1	18.15	18.15	0.05374	5.32
Reg. Cuadrática	1	7650.75	7650.75	22.652 *	5.32
Reg. Cúbica	1	0.01666	0.01666	0.000049	5.32
(Dosis)	(3)	(7668.92)			
Residuo	8	2702.00	337.75		
Total	11	10370.92			

Se verifica por la prueba de F, significancia para la regresión de grado 2.

Cálculo de las sumas de cuadrados de Regresión

(i) Obtención de los coeficientes que componen los $P_1(X_i)$, $P_2(X_i)$ y $P_3(X_i)$:

Dosis K	$y_{i.}$	$P_1(X_i)$	$P_2(X_i)$	$P_3(X_i)$
0	237	–3	1	–1
30	392	–1	–1	3
60	395	1	–1	–3
90	247	3	1	1

(ii) Obtención de las sumas de cuadrados.

$$SC\ Reg_1 = \frac{\left(\sum_{i=1}^{4} P_1(X_i)\ y_{i.}\right)^2}{r\sum_{i=1}^{4} P_1^2(X_i)} = \frac{[(-3)\,237 + (-1)\,392 + (1)\,395 + (3)\,247\,]^2}{3\,(20)} = \frac{33^2}{60} = 18.15$$

$$SC\ Reg_2 = \frac{\left(\sum_{i=1}^{4} P_2(X_i)\ y_{i.}\right)^2}{r\sum_{i=1}^{4} P_2^2(X_i)} = \frac{[(1)\,237 + (-1)\,392 + (-1)\,395 + (1)\,247\,]^2}{3\,(4)} = \frac{(-303)^2}{12} = 7650.75$$

$$SC\ Reg_3 = \frac{\left(\sum_{i=1}^{4} P_3(X_i)\ y_{i.}\right)^2}{r\sum_{i=1}^{4} P_3^2(X_i)} = \frac{[(-1)\,237 + (3)\,392 + (-3)\,395 + (1)\,247\,]^2}{3\,(20)} = \frac{1^2}{60} = 0.016666$$

Ecuación de Regresión

$$\hat{b}_0 = \bar{y}_{..} = \frac{1}{r\,t}\sum_{i=1}^{n} y_{i.} = \frac{1271}{(3)\,(4)} = 105.91667$$

$$b_1 = \frac{\sum_{i=1}^{4} P_1(X_i)\ y_{i.}}{r\sum_{i=1}^{n} P_1^2(X_i)} = \frac{33}{60} = 0.55 \quad b_2 = \frac{\sum_{i=1}^{4} P_2(X_i)\ y_{i.}}{r\sum_{i=1}^{n} P_2^2(X_i)} = \frac{-303}{60} = -25.25$$

Y también se tiene que:

$$P_1(X_i) = x_i = \frac{X_i - \overline{X}}{q} = \frac{X_i - 45}{30} = \frac{X_i}{30} - 1.5\text{, y}$$

$$P_2(X_i) = x_i^2 - \frac{n^2 - 1}{12} = \left(\frac{X_i}{30} - 1.5\right)^2 - \frac{4^2 - 1}{12} = \left(\frac{X_i}{30} - 1.5\right)^2 - 1.25.$$

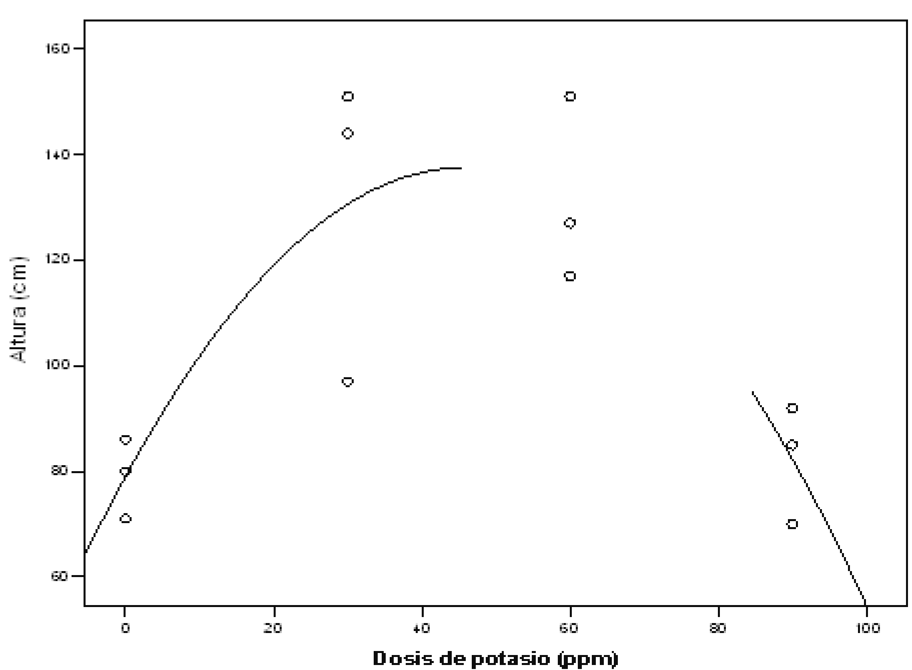
Luego:

$$\hat{y}_i = 105.91667 + 0.55\left(\frac{X_i}{30} - 1.5\right) - 2.25\left[\left(\frac{X_i}{30} - 1.5\right)^2 - 1.25\right]$$

$$\hat{y}_i = 105.91667 + 0.018333X_i - 0.825 - 2.25\left[\left(\frac{X_i^2}{900} - \frac{3X_i}{30} + 2.25\right) - 1.25\right]$$

$$\hat{y}_i = 105.91667 + 0.018333X_i - 0.825 - 0.028056X_i^2 + 2.525X_i - 25.25$$

$\hat{y}_i = 79.8417 + 2.54333X_i - 0.028056X_i^2$. Válida para valores de $X_i \in [\ 0\ ;\ 90\ ]$ ppm de K, cuyo gráfico se muestra a continuación:



De la ecuación de regresión obtenida se pueden estimar las medias de la altura asociadas a cada una de las dosis de potasio:

Dosis K	$\bar{y}_{i.}$	$\hat{y}_{i.}$
0	79.00	79.845
30	130.67	130.854
60	131.67	131.283
90	82.33	81.132

Los coeficientes de determinación (R^2) se calculan de la siguiente manera:

$$R^2 R.g.k = \frac{SC\ \mathrm{Re}\, g_k}{SC\ Dosis},\ \text{con } 0 \leq \mathrm{R}^2 \leq 1.$$

$$R^2 R.g.1 = \frac{18.15}{7668.92} = 0.0023667$$

$$R^2 R.g.2 = \frac{7650.75}{7668.92} = 0.9976$$

Determinación de la producción máxima

Para el ejemplo ilustrativo, la dosis de potasio que proporciona la altura máxima de planta, es dado por la maximización de la función de producción obtenida, es decir, a través de la aplicación del cálculo diferencial a la ecuación de regresión obtenida, o sea, derivando la función de producción: $\hat{y}_i = 79.8417 + 2.54333\mathrm{X}_i - 0.028056\mathrm{X}_i{}^2$ (2) con relación a la variable X_i, se tiene:

$$\frac{\partial\, \hat{y}_i}{\partial\, \mathrm{X}_i} = 2.54333 - 2\,(0.028056\mathrm{X}_i) \quad (3)$$

y considerando que $\frac{\partial\, \hat{y}_i}{\partial\, \mathrm{X}_i} = 0$, se obtiene: $2.54333 - 0.0562\mathrm{X}_i{*} = 0$; $2.54333 = 0.0562\mathrm{X}_i{*}$

$$\mathrm{X}_i{*} = \frac{2.54333}{0.0562} = 45.25 \approx 45\ \text{ppm}.$$

El valor de X_i* se refiere a la abscisa del punto máximo, esto es verificado cuando se deriva la ecuación (3) con relación a la variable X, y se observa que su signo es negativo, esto es:

$$\frac{\partial^2\, \hat{y}_i}{\partial\, \mathrm{X}_i{}^2} = -0.0562\mathrm{X}_i\,.$$

Así, sustituyendo el valor de X*=45 en la ecuación (2) se obtiene la altura máxima de Eucalipto, o sea:

$$\hat{y}_i = 79.8417 + 2.54333(45) - 0.028056(45)^2 \quad = \ 137.39\,\text{cm}$$

Comentarios finales

1. Gomes FP (2009) resalta que, aunque la regresión polinomial sea de gran utilidad en numerosos casos, en experimentos de fertilizantes, ella no es muy apropiada y debe, siempre que sea posible, ser reemplazada por la Ley Mitscherlich, biológicamente más aceptable y de interpretación más objetiva. Por otra parte, los polinomios son funciones poco apropiadas para representar fenómenos biológicos. Lo que los recomienda es principalmente la facilidad de su uso, pues son funciones más simples.

2. En los experimentos con fertilizantes, aunque sea significativo el componente de 3er. grado, es preferible considerar un polinomio de 2º grado, de propiedades matemáticas más acordes con el fenómeno biológico estudiado. En algunos casos se prefiere hacer uso de la regresión lineal Plateau (también conocida como regresión segmentada)

Coeficientes para interpolación de polinomios ortogonales

n = 3 niveles		n = 4 niveles			n = 5 niveles			
1er grado	2º grado	1er grado	2º grado	3er. grado	1er grado	2º grado	3er. Grado	4º Grado
–1	+1	–3	+1	–1	–2	+2	–1	+1
0	–2	–1	–1	+3	–1	–1	+2	–4
+1	+1	+1	–1	–3	0	–2	0	+6
		+3	+1	+1	+1	–1	–2	–4
					+2	+2	+1	+1

Coeficientes para más niveles, pueden ser consultados en: Gomes, FP. (2009) Curso de Estadística Experimental. Piracicaba: ESALQ p. 252 – 260

6.4 PROGRAMA EN SAS PARA EL ANÁLISIS DE REGRESIÓN POLINOMIAL EN UN EXPERIMENTO.

```
OPTIONS nodate nonumber;
DATA polin;
INPUT dos alt;
CARDS;
0	80
30	144
60	151
90	70
0	86
30	151
60	127
90	85
0	71
30	97
60	117
90	92
;
PROC print; RUN;
TITLE "Análisis de regresión polinomial";
PROC glm;
CLASS dos;
MODEL alt=dos/SS3;
CONTRAST "Reg.1" dos -3 -1  1 3;
CONTRAST "Reg.2" dos  1 -1 -1 1;
RUN;

TITLE "Modelo de regresión polinomial de grado 1";
PROC GLM DATA=polin;
MODEL alt=dos/SS1;
RUN;

TITLE "Modelo de regresión polinomial de grado 2";
PROC GLM DATA=polin;
MODEL alt=dos dos*dos/SS1;
RUN;
```

6.5 EJERCICIOS PROPUESTOS

1. Los datos siguientes se refieren a un experimento de fertilización en maíz (*Zea mays* L.), en bloques al azar. Los tratamientos constaron de fertilización con: 0, 25, 50, 75 y 100 kg/ha de P_2O_5.

Bloques	Dosis					Totales de bloques
	0	25	50	75	100	
I	3.38	7.15	10.07	9.55	9.14	39.29
II	5.77	9.78	9.73	8.95	10.17	44.40
III	4.90	9.99	7.92	10.24	9.75	42.80
IV	4.54	10.10	9.48	8.66	9.50	42.28

a) Realice el análisis de varianza
b) Ajuste un modelo de regresión polinomial.
c) Grafique la ecuación de regresión calculada en el inciso b.
d) Concluya.

2. Se realizó un experimento para conocer el efecto de la suplementación de vitamina A en la dieta de gallinas de postura, a través de la masa del huevo. Para ello se suplementaron cantidades crecientes de vitamina A en la dieta de gallinas. Se consideró como unidad experimental un grupo de 20 jaulas individuales y se realizaron cuatro repeticiones por tratamiento en un diseño completamente al azar. Además de las comparaciones entre tratamientos, el investigador deseaba conocer la dosis óptima recomendada, por lo cual se realizó un análisis de polinomios ortogonales, dado que los tratamientos de vitamina A estaban igualmente espaciados.

Repetición	Niveles de vitamina A (UI/kg)					
	2000	4000	6000	8000	10,000	12,000
1	27.9	48.3	50.8	54.5	54.2	56.6
2	32.3	50.1	52.4	52.4	51.0	50.3
3	38.6	38.4	49.6	48.0	46.5	49.0
4	34.5	50.8	51.6	51.3	52.0	50.4

Fuente: Herrera, J. G.; Barreras, A. (2001)

En este caso se tienen seis tratamientos y por lo tanto se puede ajustar un polinomio y probar los efectos lineales, cuadráticos, cúbicos, cuárticos y quíntuples.

3. Los datos que se presentan a continuación se refieren a la producción de soya (*Glycine max* L.), en kg/ha, obtenidos en el primer año de un experimento instalado con el objetivo de verificar la respuesta de la soya a la aplicación de dosis crecientes de cal. Se utilizó un diseño completamente al azar, con las siguientes dosis: 0, 2, 4, 6, 8, 10 y 12 tm/ha de cal, y seis repeticiones.

Dosis	Repeticiones					
	I	II	III	IV	V	VI
0	1458	1709	1792	2200	1075	1850
2	2559	2617	2917	2317	2259	2150
4	3050	2392	2659	2600	2959	2369
6	3059	2517	3292	2675	2834	2700
8	2842	2825	2734	3017	3000	3059
10	2300	2400	2784	2217	2409	2650
12	2375	2517	2792	2325	2504	2519

Fuente: Quaggio, J. A.; Mascarenhas, H. A. A. ; Bataglia, O. C. (1982)

Con estos datos:

a) Realice el Análisis de Varianza y calcule los polinomios ortogonales hasta el 3er. grado.

b) Obtener la ecuación de regresión polinomial más adecuada.

c) Determine la producción máxima

4. En el trabajo “Efecto de dosis de yeso en el cultivo del frijol (*Phaseolus vulgaris* L.)”, Ragazzi (1979) utilizó un diseño completamente al azar con 4 repeticiones, para estudiar los efectos de 7 dosis de yeso: 0,50,100,150,200, 250 y 300 kg/ha sobre diversas características del frijol. Para la variable peso de 1000 semillas, los resultados obtenidos, en gramos, son presentados a continuación:

Dosis	Repeticiones			
	I	II	III	IV
0	134.8	139.7	147.6	132.3
50	161.7	157.7	150.3	144.7
100	160.7	172.7	163.4	161.3
150	169.8	168.2	160.7	161.0
200	165.7	160.0	158.2	151.0
250	171.8	157.3	150.4	160.4
300	154.5	160.4	148.8	154.0

a) Realice el ANOVA

b) Efectue el procedimiento de polinomios ortogonales.

c) Emita sus conclusiones

5. En un experimento en camote (*Ipomoea batatas* L.) se aplicaron seis niveles de potasio (K_1=0, K_2=10, K_3=15, K_4=30, K_5=45, K_6=60). El objetivo fue evaluar el rendimiento (peso de tubérculos en kg/100 m^2) obtenido al aplicar los diferentes niveles de potasio. Las repeticiones fueron ubicadas en cuatro suelos con diferente fertilidad; los resultados son presentados en la tabla siguiente:

Bloques	Niveles de Potasio					
	K_1	K_2	K_3	K_4	K_5	K_6
I	40	30	29	25	40	23
II	38	36	39	31	37	25
III	35	35	25	29	38	24
IV	25	26	28	27	29	26

Con estos datos:

a) Realice el Análisis de Varianza

b) Obtenga los polinomios ortogonales hasta el 3er. grado.

c) Obtener la ecuación de regresión polinomial más adecuada.

d) Determine la producción máxima.

6. En un experimento sobre alimentación de cerdos en crecimento realizado en un diseño completamente al azar fueron utilizadas cuatro tipos de raciones: A, B, C y D. Los animales de la raza Duroc-Jersey, con edad aproximada de 3 meses. Em las raciones, harina de carne era sustituída total o parcialmente por harina de soja tostada, de tal manera que el porcentajes de esta última en las raciones eran los siguientes:

A - cero de soya (18% de harina de carne);
B - 10% de soya (12% de harina de carne);
C - 20% de soya (6% de harina de carne);
D - 30% de soya (cero de de harina de carne);

El experimento tuvo una duración de 98 días, se registraron los pesos regulares de los animales a cada 14 dias, siempre por la mañana y con animales en ayuno por más de 15 horas. En la tabla siguiente son presentados los índices de conversión (kg de ración / kg de ganancia de peso) observados durante el período de 98 días (en negrito, al final de cada columna, se presentan los totales).

Raciones			
A	**B**	**C**	**D**
3.66	3.15	3.14	3.17
3.38	3.33	3.47	3.04
2.93	3.42	3.11	2.97
3.71	3.28	3.38	3.13
3.67	3.16	3.15	2.75
3.39	3.47	3.00	2.62
3.22	3.35	3.06	3.37
3.44	2.99	3.01	3.05
27.3	**26.15**	**25.32**	**24.10**

Verifique si el efecto de las raciones es significativo y, en caso afirmativo, como los niveles de las raciones son cuantitativos, se debe descomponer los grados de libertad de las raciones en regresión de tipo lineal, cuadrática y cúbica.

****

Nota:

Aplicaciones de modelos de regresión no lineales en experimentos agronómicos, usando el programa R, puede ser consultadas en el siguiente texto, elaborado por profesores del Laboratorio de Estadística y Geoinformación, Departamento de Estadística de la Universidad Federal de Paraná (Curitiba, PR):

Zeviani, W.M.; Ribeiro Jr., P.J.; Bonat, W.H. **Curso de modelos de regressão não linear**. Presentado en la 58ª. RBRAS y 15º SEAGRO. Campina Grande, Paraíba; 22-26 de julio de 2013. 101 p.

CAPÍTULO 7
DISEÑO CUADRADO LATINO

7.1 INTRODUCCIÓN

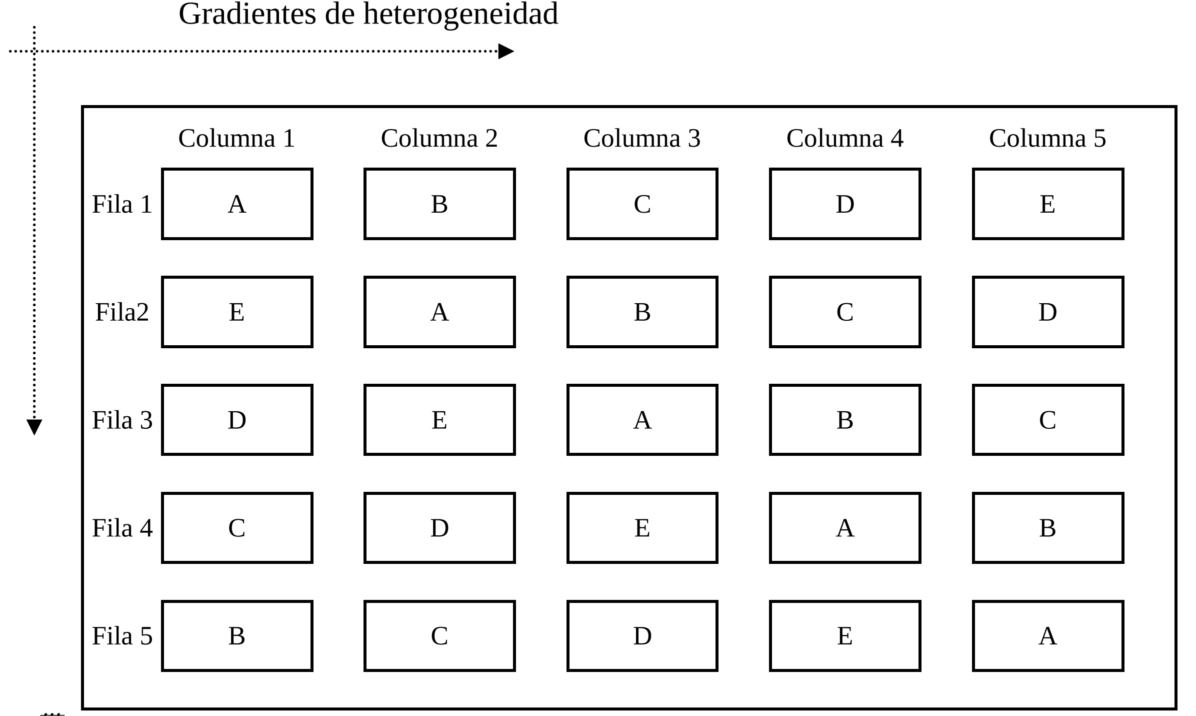
Este diseño también es conocido como diseño con un factor y dos restricciones en la aleatorización. De esta forma se tiene que el control local, representado por los bloques, es organizado de dos maneras diferentes, siendo uno organizado en el sentido de las filas (o hileras) y otro organizado en el sentido de las columnas.

En los experimentos que se realizan en el campo, este diseño es utilizado cuando existe necesidad de eliminar la heterogeneidad del suelo en dos direcciones perpendiculares, esto es, las filas en una dirección y las columnas en otra dirección.

En general, un cuadrado latino para p factores, o sea, un cuadrado latino de tamaño p×p, es un cuadrado que contiene p fila y p columnas. Cada una de las p^2 celdas contiene una de las p letras que corresponden a un tratamiento, y cada letra aparece una sola vez en cada fila y en cada columna, de tal manera que cualquier comparación de tratamientos no se vea afectada por las diferencias existentes entre hileras o entre columnas. En este diseño el número de filas o hileras (h), el número de columnas (c) y el número de tratamientos (t) debe ser igual.

7.1.1 Criterios de bloqueo

Se tiene un experimento planeado en un diseño cuadrado latino, con cinco tratamientos, identificados como: A, B, C, D y E, por lo que, se deben tener cinco filas y cinco columnas. La localización de las parcelas en el área experimental se muestra en el siguiente croquis:



Nota: Recuerde que la estructura de distribución de las unidades experimentales es conceptual y no necesariamente física. A continuación se presentan algunos ejemplos, para ilustrar esta situación.

Ejemplo 7.1

Se encuentran bajo estudio el efecto que tienen cinco reactivos distintos (A, B, C, D y E) sobre el tiempo de reacción de un proceso químico. Cada lote de material nuevo es lo suficientemente grande para permitir que sólo se realicen cinco ensayos. Mas aún, cada ensayo tarda aproximadamente una hora y media, por lo que solo pueden realizarse cinco ensayos por día. El investigador decide efectuar el experimento utilizando el diseño cuadrado latino, con el fin de controlar sistemáticamente las variables lote de material y día

La distribución de los tratamientos se muestra a continuación:

Lote	Día				
	1	2	3	4	5
1	A	B	D	C	E
2	C	E	A	D	B
3	B	A	C	E	D
4	D	C	E	B	A
5	E	D	B	A	C

Ejemplo 7.2

Un ingeniero industrial está investigando el efecto que tienen cuatro métodos de ensamblaje (A, B, C y D) sobre el tiempo de ensamblaje de un componente para televisores a color. Se seleccionaron cuatro operadores para realizar el estudio. Por otra parte, el ingeniero sabe que cada método de ensamblaje produce fatiga, por lo que el tiempo que se tarda en el último ensamblaje puede ser mayor que en el primero, independientemente del método. En otras palabras, se produce un patrón en el tiempo de ensamblaje. Para controlar esta posible fuente de variación (variabilidad), el ingeniero utilizó el diseño cuadrado latino, que aparece a continuación:

Orden de montaje	Operador			
	1	2	3	4
1	C	D	A	B
2	B	C	D	A
3	A	B	C	D
4	D	A	B	C

7.1.2 Aleatorización

Cuadrado Latino Reducido:

Es aquel en el cual la primera hilera y la primera columna están dispuestas en orden alfabético.

Procedimiento para la aleatorización

a. Seleccionar un cuadrado latino reducido

b. Permutar hileras aleatoriamente

c. Permutar columnas aleatoriamente

d. Asignar los tratamientos a las letras al azar.

Los incisos *b,c,d* se pueden realizar usando una tabla de números aleatorios, o bien a través de un simple sorteo. A continuación se muestra un ejemplo de aleatorización para un cuadrado latino de tamaño 5 × 5

a) Cuadrado latino reducido

F_1	A	B	C	D	E
F_2	B	C	D	E	A
F_3	C	D	E	A	B
F_4	D	E	A	B	C
F_5	E	A	B	C	D

b) Aleatorizar filas

	C_1	C_2	C_3	C_4	C_5
F_3	C	D	E	A	B
F_1	A	B	C	D	E
F_5	E	A	B	C	D
F_2	B	C	D	E	A
F_4	D	E	A	B	C

c) Aleatorizar columnas

C_4	C_1	C_3	C_2	C_5
A	C	E	D	B
D	A	C	B	E
C	E	B	A	D
E	B	D	C	A
B	D	A	E	C

7.2 ANÁLISIS ESTADÍSTICO

A continuación se presentan una tabla de datos para un cuadrado latino 4 x 4:

Filas	Columnas 1	2	3	4	Total filas $Y_{i..}$
1	Y_{11k}	Y_{12k}	Y_{13k}	Y_{14k}	$Y_{1..}$
2	Y_{21k}	Y_{22k}	Y_{23k}	Y_{24k}	$Y_{2..}$
3	Y_{31k}	Y_{32k}	Y_{33k}	Y_{34k}	$Y_{3..}$
4	Y_{41k}	Y_{42k}	Y_{43k}	Y_{44k}	$Y_{4..}$
Total columnas $Y_{.j.}$	$Y_{.1.}$	$Y_{.2.}$	$Y_{.3.}$	$Y_{.4.}$	$Y_{...}$ (Gran total)

siendo k = 1,2,3,4.

La tabla anterior se obtiene sin modificar el croquis de campo. Puede generarse una tabla adicional para los tratamientos, sumando todas las veces que se repitió cada tratamiento, obteniendo así el respectivo total ($Y_{..k}$), como se indica a continuación:

	Tratamientos 1	2	3	4
	Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
	Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
	Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
	Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
Total Tratamientos $Y_{..k}$	$Y_{..1}$	$Y_{..2}$	$Y_{..3}$	$Y_{..4}$
Media Tratamientos $\bar{y}_{..k}$	$\bar{y}_{..1}$	$\bar{y}_{..2}$	$\bar{y}_{..3}$	$\bar{y}_{..4}$

Media general = $\bar{y}_{...} = \frac{y_{...}}{t^2}$

7.2.1 Modelo Estadístico Matemático

$$Y_{ijk} = \mu + F_i + C_j + \tau_k + \varepsilon_{ijk}$$

i = 1,2,3, . . . t
j = 1,2,3, . . . t   h =c = t
k =1,2,3, . . . t

siendo que:

Y_{ijk} = Valor observado correspondiente al k – ésimo tratamiento en la i-ésima fila con la j-ésima columna.

μ = Media general de la variable de respuesta.

F_i = Efecto de la i - ésima fila en la variable dependiente y mide la separación de la media $\mu_{i..}$ en relación a la media μ, esto es: $F_i = \mu_{i..} - \mu$; $\mu_{i..}$= media poblacional de la i-ésima línea.

C_j = Efecto de la j - ésima columna en la variable dependiente y mide la separación de la media $\mu_{.j.}$ en relación a la media μ, esto es: $C_j = \mu_{.j.} - \mu$; $\mu_{.j.}$ = media poblacional de la j-ésima columna.

τ_k = Efecto del k - ésimo tratamiento en la variable dependiente y mide la separación de la media $\mu_{..k}$ en relación a la media μ, esto es: $\tau_k = \mu_{.j.} - \mu$; $\mu_{..k}$ = media poblacional del k-ésimo tratamiento.

ε_{ijk} = Error experimental asociado al valor observado Y_{ijk}.

7.2.2 Supuestos

$\varepsilon_{ijk} \sim NID(0, \sigma^2)$

No existe interacción entre filas y tratamientos.

No existe interacción entre columnas y tratamientos.

7.2.3 Hipótesis

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)

Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (Al menos uno de los tratamientos produce efectos distintos)

7.2.4 Tabla de Análisis de Varianza

FV	GL	SC	CM	Valor de F
Tratamientos	t – 1	$\frac{\sum_{k=1}^{t} Y_{..k}^2}{t} - \frac{Y_{...}^2}{t^2}$	SC_{trat}/GL_{trat}	CM_{trat}/CM_{ee}
Filas	t – 1	$\frac{\sum_{i=1}^{t} Y_{i..}^2}{t} - \frac{Y_{...}^2}{t^2}$		
Columnas	t – 1	$\frac{\sum_{j=1}^{t} Y_{.j.}^2}{t} - \frac{Y_{...}^2}{t^2}$		
Error experimental	(t –1) (t – 2)	SC total– (SC trat + SCH + SCC)	SC_{ee}/GL_{ee}	
Total	$t^2 - 1$	$\sum_{i=1}^{t} \sum_{j=1}^{t} \sum_{k=1}^{t} Y_{ijk}^2 - \frac{Y_{...}^2}{t^2}$		

Regla de Decisión

Rechazar Ho. sí Valor de F ≥ Valor crítico de F $_{(gl\ trat;\ gl\ error;\ \alpha)}$
No Rechazar Ho. sí Valor de F < Valor crítico de F $_{(gl\ trat;\ gl\ error;\ \alpha)}$

7.2.5 Ejemplo de aplicación

Gomes, FP (2009) reporta un experimento sobre evaluación de variedades de caña de azúcar realizado en Brasil, en el que fueron usadas cinco variedades: Co-290 (A), Co-421 (B), Co-419 (C), POJ-2878 (D) y CP-3613 (E), dispuestas en un cuadrado latino de tamaño 5×5. Las producciones de caña en kilogramos por parcela son dadas en la tabla siguiente:

						$Y_{i..}$
	D 432	A 518	B 458	C 583	E 331	**2322**
	C 724	E 478	A 524	B 550	D 400	**2676**
	E 489	B 384	C 556	D 297	A 420	**2146**
	B 494	D 500	E 313	A 486	C 501	**2294**
	A 515	C 660	D 438	E 394	B 318	**2325**
$Y_{.j.}$	**2654**	**2540**	**2289**	**2310**	**1970**	**11763**

Cuadro Auxiliar

Variedad	Total Tratamientos $Y_{..k}$	Media Tratamientos $\bar{Y}_{..k}$
Co-290 (A)	2463	492.6
Co-421 (B)	2204	440.8
Co-419 (C)	3024	604.8
POJ-2878 (D)	2067	413.4
CP-3613 (E)	2005	401.0
	11763	2352.6

$$SCtrats = \frac{(2,463)^2 + \ldots + (2,005)^2}{5} - \frac{11,763^2}{25} = 137,488$$

$$SCfilas = \frac{(2,322)^2 + \ldots + (2,325)^2}{5} - \frac{11,763^2}{25} = 30,480$$

$$SCcolumnas = \frac{(2,654)^2 + \ldots + (1,970)^2}{5} - \frac{11,763^2}{25} = 55,640$$

$$SCtotal = (432)^2 + (518)^2 + \ldots + (318)^2 - \frac{11,763^2}{25} = 257,724$$

$$SCerror = 257,724 - 137,488 - 30,480 - 55,640 = 34,116$$

Resumen del análisis de varianza

Fuentes de variación	Grados de libertad	Suma de cuadrados	Cuadrados Medios	Valor de F	F crítica (4,12,0.05)
Variedades	4	137,488	34,372	12.09*	3.26
Filas	4	30,480			
Columnas	4	55,640			
Error Experimental	12	34,116	2,843		
Total	24	257,724			

$$CV = \frac{\sqrt{CME}}{\bar{Y}_{...}} \times 100 = \frac{\sqrt{2,843}}{470.52} \times 100 = 11.33\%$$

De acuerdo con los resultados del análisis de varianza, se concluye que las variedades presentan efectos diferenciados en cuanto a la producción de caña de azúcar, por lo tanto se recomienda efectuar el respectivo análisis post-anova.

Prueba de comparación múltiple de medias de acuerdo al criterio de Tukey

1. Matriz de diferencias

Variedad	Media	Co-419 604.80	Co-290 492.60	Co-421 440.80	POJ-2878 413.40	CP-3613 401.00
CP-3613	401.00	203.80 *	91.60	39.80	12.40	–
POJ-2878	413.40	191.40 *	79.20	27.40	–	
Co-421	440.80	164.0 *	51.80	–		
Co-290	492.60	112.20 *	–			
Co-419	604.80	–				

2. Cálculo del comparador W

$$W = q_{(t,glee,\alpha)} \times \sqrt{\frac{CMee}{t}} \qquad W = 4.51 \times \sqrt{\frac{2843}{5}} = 107.54$$

3. Presentación de los resultados

Variedad	Media	Grupo Tukey
Co-419	604.80	a
Co-290	492.60	b
Co-421	440.80	b
POJ-2878	413.40	b
CP-3613	401.00	b

7.3 DISEÑO CUADRADO LATINO (DCL) CON DATOS FALTANTES

7.3.1 Un dato faltante

Cuando se tiene una unidad experimental perdida se utiliza la ecuación que se presenta a continuación:

$$Y_{ijk} = \frac{t\,(H + C + T) - 2S}{(t-1)\,(t-2)}$$

Siendo:

t = número de tratamientos

H = total de valores observados para la hilera que contiene la unidad faltante

C = total de valores observados para la columna que contiene la unidad faltante

T = total del tratamiento que contiene la unidad faltante

S = gran total de valores observados.

Luego se efectúa el ANOVA, restando uno (1) a los grados de libertad del total y del error experimental. Cuando se efectúan las comparaciones de medias de tratamientos se deben observar las mismas precauciones mencionadas para el diseño en bloques al azar; es decir, que cuando se esté comparando el tratamiento al cual le estimamos el valor se deberá usar el error estándar siguiente:

$$S\overline{x} = \sqrt{\frac{1}{2} CME \left[\frac{2}{t} + \frac{t}{(t-1)\,(t-2)} \right]}$$

7.3.2 Dos datos faltantes

Si los valores perdidos son dos o más usar el método iterativo (descrito en bloques al azar) empleando las siguientes ecuaciones:

1. Para calcular el primero de los datos faltantes

$$a = \frac{\bar{Y}_{i,,} + \bar{Y}_{.j.} + \bar{Y}_{..k}}{3}$$

2. Para calcular el segundo dato faltante y continuar con el proceso iterativo:

$$Y_{ijk} = \frac{t\,(H + C + T) - 2S}{(t-1)\,(t-2)}$$

Luego se efectúa el ANOVA, restando dos (2) a los grados de libertad del total y del error experimental.

7.3.3 Ejercicio de aplicación

Los datos a seguir se refieren a la producción de yuca (kg) en un experimento citado por Nogueira (1997) que involucró cuatro sistemas de plantación de estacas de yuca, conducido mediante un diseño cuadrado latino de tamaño 4×4. Los tratamientos evaluados fueron:

A. Estacas con 0.30 m de largo plantadas con el sistema tradicional (o común).

B. Estacas con 0.30 m de largo plantadas con 0.15 m enterradas e inclinadas.

C. Estacas con 0.30 m de largo plantadas con 0.15 m enterradas, inclinadas y en camellón.

D. Estacas con 0.30 m de largo plantadas en forma horizontal en la superficie del camellón.

Filas	Columnas			
	1	2	3	4
1	**A** X	**D** 98.8	**B** 122.6	**C** 102.5
2	**B** 126.3	**A** 110.3	**C** 110.1	**D** 53.7
3	**D** 83.1	**C** 106.4	**A** 100.6	**B** 93.4
4	**E** 96.7	**B** 107.2	**D** 75.7	**A** 80.2

a) Estime el dato correspondiente al tratamiento A en la fila 1 y columna 1 y realice el análisis de varianza.

b) Con los siguientes resultados, estime los datos faltantes (tratamiento A en la fila 1 y columna 1 y el tratamiento C en la fila 2 y columna 3) y realice el análisis de varianza

Filas	Columnas			
	1	2	3	4
1	**A** X	**D** 98.8	**B** 122.6	**C** 102.5
2	**B** 126.3	**A** 110.3	**C** X	**D** 53.7
3	**D** 83.1	**C** 106.4	**A** 100.6	**B** 93.4
4	**E** 96.7	**B** 107.2	**D** 75.7	**A** 80.2

7.4 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO CUADRADO LATINO

```
OPTIONS nodate nonumber;
DATA dcl;
INPUT var $ fil col prod;
CARDS;
D   1   1   432
C   2   1   724
E   3   1   489
B   4   1   494
A   5   1   515
A   1   2   518
E   2   2   478
B   3   2   384
D   4   2   500
C   5   2   660
B   1   3   458
A   2   3   524
C   3   3   556
E   4   3   313
D   5   3   438
C   1   4   583
B   2   4   550
D   3   4   297
A   4   4   486
E   5   4   394
E   1   5   331
D   2   5   400
A   3   5   420
C   4   5   501
B   5   5   318
;
PROC anova;
   CLASS var fil col;
   MODEL prod= var fil col;
   MEANS var/TUKEY; RUN;
```

7.5 DISEÑO CUADRADO LATINO (DCL) CON MUESTREO

En un experimento con la variedad de tomate Imperial Rojo (ejemplo hipotético), se evaluaron cuatro distanciamientos de siembra:

A.	0.6 metros	C.	1.0 metros
B.	0.8 metros	D.	1.2 metros,

determinando su efecto sobre el rendimiento de frutos (expresado en kilogramos por parcela neta). Los resultados se muestran en el siguiente croquis de campo:

Filas (i)	Columnas (j)				$Y_{i...}$
	1	2	3	4	
	A 15 14	**B** 18 17	**D** 15 15	**C** 14 13	
$Y_{ijk.}$	29	35	30	27	121
	C 19 15	**D** 13 13	**B** 12 16	**A** 17 16	
$Y_{ijk.}$	34	26	28	33	121
	D 12 14	**A** 16 15	**C** 18 16	**B** 16 15	
$Y_{ijk.}$	26	31	34	31	122
	B 17 13	**C** 20 16	**A** 13 15	**D** 13 12	
$Y_{ijk.}$	30	36	28	25	119
$Y_{.j..}$	119	128	120	116	483

Fuente: datos hipotéticos.

Cuadro auxiliar

Distanciamiento	Total Tratamientos $Y_{..k}$	Media Tratamientos $\bar{Y}_{..k}$
(A)	121	15.13
(B)	124	15.50
(C)	131	16.38
(D)	107	13.38

El modelo estadístico matemático asociado a este diseño es:

$$Y_{ijkl} = \mu + F_i + C_j + \tau_k + \varepsilon_{ijk} + \eta_{ijkl}$$

i = 1,2,3, . . . t
j = 1,2,3, . . . t
k =1,2,3, . . . t
l = 1,2,3,. . . m (muestras)
h =c = t

siendo que:

Y_{ijkl} = valor de la variable de respuesta correspondiente a la l-ésima muestra sobre la unidad experimental que lleva el tratamiento k en la fila i y columna j.

μ = Media general de la variable de respuesta.

F_i = Efecto de la i - ésima fila

C_j = Efecto de la j - ésima columna

τ_k = Efecto del k - ésimo tratamiento

η_{ijkl} = Error de muestreo dentro de la ijk-ésima unidad experimental.

ε_{ijk} = Error experimental asociado a la ijk - ésima unidad experimental

Supuestos

$\varepsilon_{ijk} \sim NID\ (0, \sigma^2)$

No existe interacción entre filas y tratamientos.

No existe interacción entre columnas y tratamientos.

$\eta_{ijkl} \sim NID\ (0, \sigma_m^2)$

Hipótesis

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)

Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (Al menos uno de los tratamientos produce efectos distintos)

Análisis de Varianza

El cuadro resumen del análisis de varianza se presenta en la tabla siguiente:

FV	GL	SC	CM
Tratamientos	t – 1	$\frac{\sum_{k=1}^{t} Y_{..k.}^2}{tm} - \frac{Y_{....}^2}{t^2 m}$	SC_{trat}/GL_{trat}
Filas	t – 1	$\frac{\sum_{i=1}^{t} Y_{i...}^2}{tm} - \frac{Y_{....}^2}{t^2 m}$	

Columnas	t – 1	$\frac{\sum_{j=1}^{t} Y_{.j..}^2}{tm} - \frac{Y_{....}^2}{t^2 m}$	
Error experimental	(t –1) (t – 2)	$SC_{total} - Sc_{em} - SC_{col} - SC_{filas} - SC_{trat}$	SCee/Glee
Error de muestreo	t^2 (m – 1)	$\sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t}\sum_{l=1}^{m} Y_{ijkl}^2 - \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t} \frac{Y_{ijk.}^2}{m}$	SCem/GLem
Total	t^2m – 1	$\sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t} Y_{ijkl}^2 - \frac{Y_{....}^2}{t^2 m}$	

$$SCtrats = \frac{(121)^2 + (124)^2 + (131)^2 + (107)^2}{(4)(2)} - \frac{483^2}{(4)^2(2)} = 38.09$$

$$SCfilas = \frac{(121)^2 + (121)^2 + (122)^2 + (119)^2}{(4)(2)} - \frac{483^2}{32} = 0.59$$

$$SCcolumnas = \frac{(119)^2 + (128)^2 + (120)^2 + (116)^2}{(4)(2)} - \frac{483^2}{32} = 9.84$$

$$SCtotal = (15)^2 + (14)^2 + \ldots + (12)^2 - \frac{483^2}{32} = 7421 - 7290.28 = 130.72$$

$$SCem = 7421 - \left(\frac{29^2 + 35^2 + \ldots + 25^2}{2}\right) = 41.50$$

$$SCee = 130.72 - (0.59 + 9.84 + 38.09 + 41.50) = 40.70$$

El resumen del análisis de varianza se presenta en el siguiente cuadro:

FV	GL	SC	CM	Valor de F
Tratamientos	3	38.09	12.70	$F_2 = 3.40$ *
Filas	3	0.59		
Columnas	3	9.84		
Error experimental	6	40.70	6.78	$F_1 = 2.62$ NS
Error de muestreo	16	41.50	2.59	
Total	31	130.72		

$$F_1 = \frac{CMee}{CMem} = \frac{6.78}{2.59} = 2.62 \qquad F_{critica(6,16,0.05)} = 2.74$$

Conclusión:

Como el valor de la estadística F_1 es menor que el valor crítico de F, entonces, se procede a calcular el valor de la estadística F_2, con el valor de cuadrado medio de tratamientos (CM_{trat}) y el cuadrado medio del error ponderado (CM_{ep}) de la siguiente forma:

$$CMep = \frac{SCee + SCem}{glee + glem} = \frac{40.70 + 41.50}{6 + 16} = 3.74$$

$F_2 = \frac{CMtrat}{CMep} = \frac{12.70}{3.74} = 3.40$, que se compara con el valor crítico de F que está en función de los grados de libertad de tratamientos, grados de libertad del error ponderado y un determinado nivel de significancia y cuyo valor es $F_{crítica(3,22,0.05)} = 3.05$

Como $F_2 \geq F_{crítica(3,22,0.05)}$, se concluye que los distanciamientos de siembra evaluados producen diferentes efectos en la producción de frutos de tomate. Por lo que se recomienda efectuar un análisis post-andeva.

7.6 PROGRAMA EN SAS PARA EL ANÁLISIS DE UN EXPERIMENTO EN UN DISEÑO CUADRADO LATINO CON MUESTREO

```
OPTIONS nodate nonumber;
DATA dclsub;
INPUT trat $ fil col rend;
CARDS;
A    1    1    15
A    1    1    14
C    2    1    19
C    2    1    15
D    3    1    12
D    3    1    14
B    4    1    17
B    4    1    13
B    1    2    18
B    1    2    17
D    2    2    13
D    2    2    13
A    3    2    16
A    3    2    15
C    4    2    20
C    4    2    16
D    1    3    15
D    1    3    15
B    2    3    12
B    2    3    16
C    3    3    18
C    3    3    16
A    4    3    13
A    4    3    15
C    1    4    14
C    1    4    13
A    2    4    17
A    2    4    16
B    3    4    16
B    3    4    15
D    4    4    13
D    4    4    12
;
PROC anova;
TITLE "DCL con muestreo";
CLASS trat fil col;/*con submuestreo*/
MODEL rend=trat fil col trat(fil*col);/*trat(fil*col)=error experimental*/
TEST h=trat e=trat(fil*col);
RUN;
PROC anova;
TITLE "mancomunando errores"; /*sin efecto de muestreo*/
CLASS trat fil col;
MODEL rend = trat fil col;
RUN;
```

En Infostat® puede resolverese el ejercicio anterior, para ello siga las siguientes instrucciones:

1. Ingrese los datos de la misma forma como están descritos en el programa de SAS (recuerde etiquetar las columnas).

2. En el menú ESTADÍSTICAS ----→ ANÁLISIS DE VARIANZA, digite en VARIABLES DEPENDIENTES: rend ; y en VARIABLES DE CLASIFICACIÓN: trat, fil y col.

3. Luego en ESPECIFICACIÓN DE LOS TÉRMINOS DEL MODELO, digite: trat , fil, col, fil*col. El programa generará la siguiente salida:

F.V.	SC	gl	CM	F	p-valor
Trat	38.09	3	12.70		
fil	0.59	3	0.20		
col	9.84	3	3.28		
fil*col	40.69	6	6.78	2.61	0.0582 NS
Error	41.50	16	2.59		
Total	130.72	31			

Nota: "fil*col" representa el error experimental y "Error" el error de muestreo.

4. Luego, tendrá las siguientes opciones:

4.1 Como el error de muestreo no fue significativo, se procede a calcular el valor de la estadística F2, con el valor de cuadrado medio de tratamientos (CMtrat) y el cuadrado medio del error ponderado (CMep). Para ello retire el efecto de "fil*col" en la especificación del modelo y ejecute de nuevo el programa. Los resultados serán los siguientes:

F.V.	SC	gl	CM	F	p-valor
Trat	38.09	3	12.70	3.40	0.0358*
fil	0.59	3	0.20		
col	9.84	3	3.28		
Error	82.19	22	3.74		
Total	130.72	31			

4.2 En aquellos casos en que el error de muestreo es significativo, en la especificación del modelo digite: Trat\fil*col, fil, col, fil*col. Infostat realizará el cálculo de la estadística F de tratamientos, considerando el CM del error experimental (6.78). El resultado será:

F.V.	SC	gl	CM	F	p-valor	(Error)
Trat	38.09	3	12.70	1.87	0.2351	(fil*col)
fil	0.59	3	0.20			
col	9.84	3	3.28			
fil*col	40.69	6	6.78	2.61	0.0582	
Error	41.50	16	2.59			
Total	130.72	31				

Compare los resultados obtenidos con SAS e Infostat y el cálculo manual. Como tarea, realice el análisis posterior al ANOVA, considerando el término de error apropiado.

7.7 EJERCICIOS PROPUESTOS

1. Los siguientes datos se refieren al promedio de producción de leche durante un mes de experimento, de vacas Holstein altas productoras, las cuales fueron alimentadas con una dieta en la cual se varió la cantidad de niacina (0, 4, 8, 12 y 16 g/kg de materia seca). Los animales fueron agrupados por edades (filas) y condición corporal (columnas) al inicio del experimento, la cual fue evaluada en escala de 1 a 10.

   Edades: 26, 40, 43, 58 y 60 meses de edad.
   Condición corporal: >9, 8, 7, 6 y 5.
   Tratamientos: A = 0, B = 4, C= 8, D= 12 y E= 16 g/kg de MS

   La producción de leche agrupada por filas, columnas y tratamientos se muestra a continuación:

	Columnas				
	(D) 732	(E) 854	(C) 641	(B) 610	(A) 549
	(A) 728	(B) 730	(D) 854	(C) 762	(E) 976
Filas	(E) 1010	(A) 750	(B) 860	(D) 720	(C) 1000
	(C) 900	(D) 1100	(A) 860	(E) 1200	(B) 920
	(B) 980	(C) 970	(E) 1250	(A) 930	(D) 1070

Fuente: Herrera, J.; Barrera, A. (2001)

   Realice el análisis de varianza, incluyendo una prueba de Tukey si fuese necesario. Indique sus conclusiones claramente.

2. Se probaron 4 raciones alimenticias para pollos, criados en jaula tipo batería de 4 pisos (filas) y 4 casilleros (columnas). La variable analizada fue: peso del pollo (kg.) a las 8 semanas de edad

	Casilleros			
Pisos	**1**	**2**	**3**	**4**
1	1.40(A)	1.38(B)	1.40(C)	1.60(D)
2	1.35(B)	1.28(A)	1.45(D)	1.62(C)
3	1.38(C)	1.40(D)	1.42(B)	1.63(A)
4	1.39(D)	1.39(C)	1.40(A)	1.60(B)

a) Presente el Modelo Aditivo Lineal
b) Realice la Prueba de Hipótesis correspondiente. Use α=0.05
c) Realice la Prueba de Tukey para comparar si existe diferencias entre los tratamientos en estudio. Use α=0.05

3. En un ensayo de alimentación de cerdos reportado por Gomes, FP (2009), se usó un diseño cuadrado latino 4 x 4, con los resultados siguientes, referentes a la ganancia en peso, expresada en kilogramos, al final de 252 días.

Filas	**Columnas**			
	1	**2**	**3**	**4**
Camada 1	A=93.0	B=108.6	C=108.9	D=102.0
Camada 2	B=115.4	D=96.5	A =77.9	C=100.2
Camada 3	C=102.1	A=94.9	D=116.9	B=96.0
Camada 4	D=117.6	C=114.1	B=118.7	A=97.6

Los tratamientos fueron: **A**: Castración a los 56 días de edad; **B**: Animales enteros; **C**: Castración a los 7 días de edad; **D**: Castración a los 21 días. Las columnas tenían como objetivo controlar la variación de peso de los lechones dentro de cada camada. Con esta información analice los datos y compare las medias de los tratamientos.

4. Veras, E. (1992) realizó el trabajo de tesis de grado "Evaluación de dos atrayentes sexuales y tres mezclas de éstos en captura de mosca de la fruta del mediterráneo (*Ceratitis capitata* Wied). Los tratamientos evaluados fueron:

Tratamiento	Descripción
T1	Trimedlure
T2	Capilure
T3	50% Capilure + 50% Trimedlure
T4	75% Capilure + 25% Trimedlure
T5	25% Capilure + 75% Trimedlure

La variable de respuesta evaluada fue: número de moscas atrapadas por tratamiento. Los resultados para la semana 16 se presentan a continuación:

Fila	Columna	Tratamiento	Número de moscas atrapadas.
1	1	5	0
1	2	3	65
1	3	2	102
1	4	1	0
1	5	4	28
2	1	4	55
2	2	5	0
2	3	3	27
2	4	2	55
2	5	1	0
3	1	1	0
3	2	4	99
3	3	5	9
3	4	3	44
3	5	2	17
4	1	2	19
4	2	1	2
4	3	4	55
4	4	5	2
4	5	3	12
5	1	3	16
5	2	2	53
5	3	1	1
5	4	4	4
5	5	5	1

a) Verifique los supuestos de normalidad y de homogeneidad de varianzas; y en caso de ser necesario, realice la transformación más adecuada.
b) Realice el ANOVA y concluya.
c) De ser necesario aplique una prueba de comparación múltiple de medias.

5. Un experimento de aclareos del área basal fue realizado en el ejido La Victoria, Pueblo Nuevo, Durango en un bosque de *Pinus cooperi*. Se usó un cuadro latino de 6 × 6 para evaluar 6 intensidades (en %) de aclareo del área basal y estudiar la influencia que esta práctica silvícola ejerce sobre el diámetro, la altura, la proyección de copa, etc, del arbolado. Como variable respuesta, se midieron los incrementos del diámetro normal (a 1.3 m) en un periodo de cuatro años, y se consideran 3 submuestras por unidad experimental. El % de aclareo del área basal en los seis tratamientos aplicados en este experimento, fueron los siguientes: A = 20%, B = 30%, C = 50%, D = 70% , E = 100% y F= 0%. Los tratamientos se distribuyeron así sobre el terreno:

	Columnas					
Filas	1	2	3	4	5	6
1	F	E	D	C	B	A
2	E	C	A	D	F	B
3	B	A	F	E	D	C
4	A	B	E	F	C	D
5	D	F	C	B	A	E
6	C	D	B	A	E	F

y los datos usados para en este ejemplo son los siguientes:

	Columnas					
Filas	1	2	3	4	5	6
	1.6	1.5	3	2.9	1.9	1.5
1	2.2	2.1	1.8	1.8	1.6	3
	2.7	3.4	3.1	3.7	2.8	3.3
	1.7	3.1	2.4	3	1.4	1.8
2	3.4	2.2	2.5	1.7	3	1.7
	2	2.7	2.5	1.8	2.8	2.4
	1.8	2.3	2.8	1.7	1.8	2.2
3	2.7	2	2.2	3.1	3.3	1.3
	2.6	3.5	1.8	3.5	2.4	1.4
	1.9	1.8	2.5	3	1.8	1.5
4	2.6	2.9	2.5	1.9	2.7	2.9
	1.7	1.7	2.6	2.4	2.3	2
	2.6	3	1.6	2.3	3	3.2
5	1.8	2.2	2.3	2.2	1.9	3.3
	1.9	2.2	2.3	3	2.3	2.5
	1.8	2.4	2.2	2	2.6	3.3
6	2.8	1.8	3	1.9	3.2	2.4
	2.3	2.3	1.5	2.2	2.2	2.6

Realice el ANOVA de un experimento en cuadro latino con submuestreo.

6. Un experimento para evaluar 4 variedades de frijol (A, B, C y D) fue realizado en invernadero, utilizando un diseño cuadrado latino, considerando lecturas de 3 plantas por cada unidad experimental. La variable respuesta medida fue el peso de la masa de materia seca (gramos) de cada una de las plantas a los 30 días después de la siembra. Los resultados se presentan a continuación (los totales por unidad experimental se presentan en negrito junto al nombre del tratamiento)

Columnas

Filas	**1**	**2**	**3**	**4**
1	6.3 7.4 6.3 **20 A**	6.4 7.8 6.8 21 **B**	6.5 8.2 7.5 **22.2 C**	6.8 9 8.1 **23.9 D**
2	6.1 7 8.9 **22 D**	6.5 7.2 7.5 **21.2 A**	6.6 7.3 7.9 **21.8 B**	6.9 7.4 8.4 **22.7 C**
3	6.4 7.8 7.5 **21.7 C**	6.9 7.9 8 **22.8 D**	6 7.1 6 **19.1 A**	6.2 7.5 6.2 **19.9 B**
4	6.7 8.2 9.4 **24.3 B**	6.8 8.4 9.9 **25.1 C**	6.9 8.5 10.5 **25.9 D**	6 8 8 **22 A**

Fuente: Quiroga, V. 1976. Manual práctico para el análisis de experimentos. San José, Costa Rica: IICA, Programa de Información Agropecuaria del Istmo Centroamericano. 113 p.

a) Describa el modelo estadístico matemático asociado a este experimento
b) Realice el análisis de varianza. Verifique la significancia del error de muestreo.
c) En caso de ser necesario realice la prueba de comparación múltiple de medias.

CAPÍTULO 8
EXPERIMENTOS FACTORIALES

8.1 INTRODUCCIÓN

Experimentos factoriales son aquellos en los que se estudia simultáneamente el efecto de dos o más factores, y donde los tratamientos se forman por la combinación de los diferentes niveles de cada uno de los factores. Los experimentos factoriales en si no constituyen un diseño experimental si no que ellos deben ser llevados en cualquiera de los diseños tal como: completamente al azar, bloques al azar, cuadrado latino, etc.

Pueden considerarse dos tipos fundamentales de experimentos factoriales: a) *factorial completo*, el cual ensaya todas las posibles combinaciones de tratamientos que se generan con los distintos niveles de los factores en estudio; b) *factorial fraccionado*, el cual ensaya solo alguna de las posibles combinaciones de tratamientos que pueden generarse.

Dentro de los factoriales completos pueden distinguirse los factoriales *simétricos* en donde cada factor presenta igual número de niveles, pudiéndose representar mediante la notación $\boldsymbol{p^n}$, siendo p el número de niveles de los factores y n el número de factores, ejemplo: 3^2 indica que tenemos 2 factores con 3 niveles cada uno, y en total tendremos 9 tratamientos. Y los asimétricos en donde cada factor presenta desigual número de niveles, pudiendo representarse como $\boldsymbol{p} \times \boldsymbol{q}$, siendo p y q los niveles del factor A y del factor B, respectivamente.

8.1.1 Conceptos fundamentales

a) Factor: Un factor es cada una de las variables independientes, cuyo efecto se está interesado en evaluar. Generalmente se denotan con letras mayúsculas (A,B, . . . Z o con las iniciales de los factores a probar). Los factores pueden ser cuantitativos (cantidad de fertilizante, de insecticida, una hormona, de tiempo, temperatura, concentración, etc.) o cualitativos (variedades, métodos de aplicación, marcas de producto, razas, procedencia, etc.).

b) Nivel: Un nivel de un factor es un cada uno de los valores o modalidades que constituyen un factor. Si el factor es cuantitativo, los niveles están formados por las cantidades o dosis del mismo; si el factor es cualitativo los niveles los constituyen las manifestaciones del mismo o los tratamientos dentro del factor (nombres de las variedades, diferentes métodos de aplicación, las diferentes marcas del producto, las diferentes razas, etc.). Por lo general se identifican con letras minúsculas (a_1, a_2, . . . ; b_1, b_2, . . .; v_1, v_2, . . .; etc.)

c) Efecto de un factor: El efecto de un factor se define como el cambio en la respuesta producida por un cambio en el nivel del factor. Con frecuencia, éste se conoce como: EFECTO PRINCIPAL, porque se refiere a los factores de interés primordial del experimento. Por ejemplo, considere los datos de la siguiente tabla, referentes a un experimento factorial 2 × 2.

		Factor B	
	Niveles	B_1	B_2
Factor A	A_1	20	30
	A_2	40	52

El efecto principal del factor A podría interpretarse como la diferencia entre la respuesta promedio en el primero y segundo nivel de ese factor. Numéricamente,

$$A = \frac{40+52}{2} - \frac{20+30}{2} = 21$$

Esto es, incrementar el factor A del nivel 1 al nivel 2 produce un cambio en la respuesta promedio de 21 unidades. De forma análoga, el efecto principal de B es:

$$B = \frac{30+52}{2} - \frac{20+40}{2} = 11$$

d) Efecto de la interacción: en algunos experimentos puede encontrarse que la diferencia en la respuesta entre los niveles de un factor no es la misma en todos los niveles de los otros factores. Cuando esto ocurre existe una interacción entre los factores. Por ejemplo, los siguientes datos referentes a un experimento factorial 2^2 con interacción:

		Factor B	
	Niveles	B_1	B_2
Factor A	A_1	20	40
	A_2	50	12

En el primer nivel del factor B, el efecto de A es: A = 50 – 20 = 30, mientras que en el segundo nivel de B, el efecto de A es: A = 12 – 40 = –28.

Puede observarse que existe una interacción entre los factores A y B porque el efecto de A depende del nivel seleccionado de B. Estas ideas pueden ilustrarse gráficamente. En la Figura 1 se muestra una gráfica de la respuesta de los datos de la Tabla 1 contra los niveles del factor A para ambos niveles del factor B.

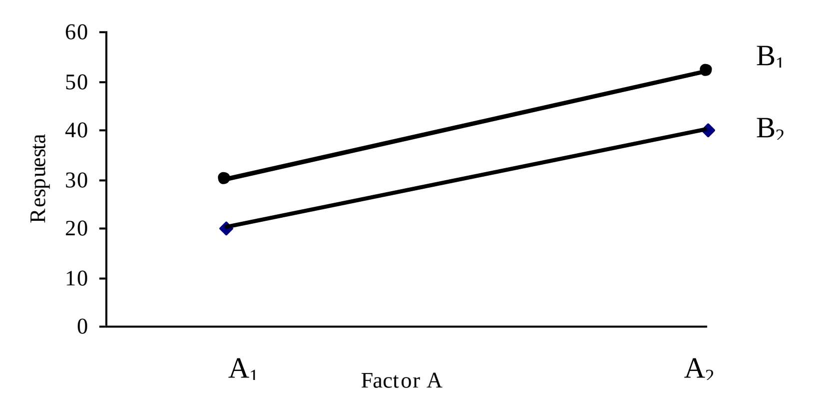
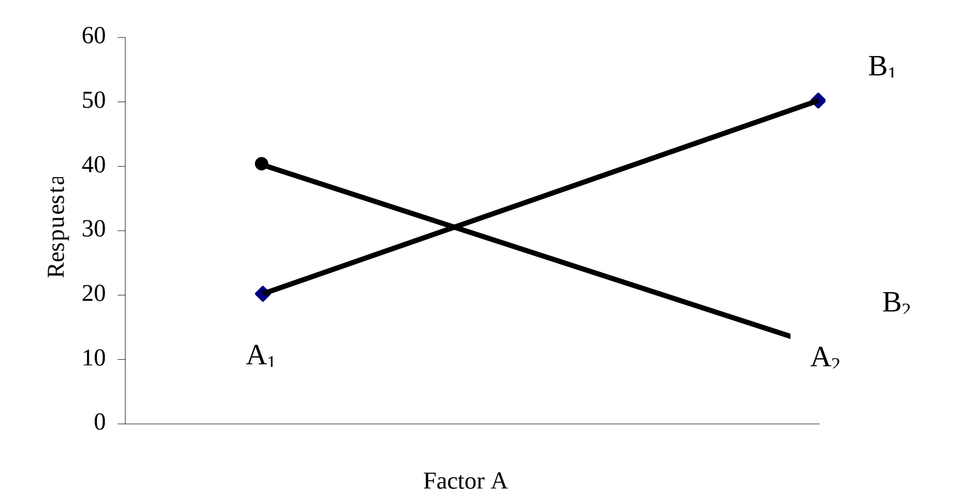


Figura 1 Ejemplo de un experimento bifactorial sin interacción

Note que las rectas B_1 y B_2 son aproximadamente paralelas. Esto indica que no hay interacción entre los factores A y B.

De igual forma, en la Figura 2 se presenta una gráfica de la respuesta de los datos de la Tabla 2. En este caso se ve que las rectas B_1 y B_2 no son paralelas. Esto muestra que existe una interacción entre A y B.



Con frecuencia estas gráficas son muy útiles para interpretar interacciones significativas, sin embargo, no debe ser la única técnica para analizar los datos, porque su interpretación es subjetiva y su apariencia, a menudo, es engañosa.

8.1.2 Ventajas

a. Se logra una gran eficiencia en el uso de los recursos experimentales disponibles, al estudiar dos o más factores en un único experimento.

b. Los resultados experimentales son aplicables a un rango de condiciones más amplio debido a las combinaciones de los diversos factores en un solo experimento. Los resultados son de naturaleza más comprensiva.

c. Se obtiene información respecto a las diversas interacciones. Por medio de los efectos de las interacciones se verifica si un factor es independiente o dependiente del o los otro(s).

Si la prueba de F de una interacción da resultados no significativos, concluimos que los factores son independientes, esto es, el comportamiento de un factor independe de la variación (ausencia o presencia) del otro factor. Las conclusiones por separado para uno y otro factor son válidas.

Por otra parte, si la prueba de F de una interacción da significativa, entonces concluimos que la respuesta de un factor es dependiente de la presencia o ausencia del otro factor, o sea que ellos son dependientes. En esas condiciones, debemos estudiar el comportamiento de un factor, dentro de cada nivel del otro factor. Por ejemplo, estudiaríamos (si la interacción $A \times B$ es significativa):

Entre A dentro de B_1, o alternativamente, Entre B dentro de A_1

Entre A dentro de B_2 Entre B dentro de a_2

… …

Entre A dentro de B_j Entre B dentro de A_i,

Si pretendemos estudiar el comportamiento de 4 cultivares de maíz en 3 distanciamientos en 5 bloques; el cálculo de los grados de libertad queda de la siguiente manera:

Fuentes de Variación	Grados de libertad
Cultivares (C)	a – 1
Distanciamientos (D)	b – 1
Interacción C x D	(a–1) (b–1)
(Tratamentos)	(ab–1)
Bloques	r – 1
Resíduo	(ab – 1) (r –1)
Total	abr –1

La parte inferior de ese cuadro es un análisis en bloques al azar con ab tratamientos y r bloques. Este análisis es llamado análisis preliminar. La parte superior es análisis en arreglo combinatorio que corresponde a la descomposición de los (ab – 1) grados de libertad de los tratamientos.

Considerando que a = 4 cultivares, b = 3 distanciamientos y r = 5 bloques, la tabla con grados de libertad queda de la siguiente manera:

Fuentes de Variación	Grados de libertad
Cultivares (C)	3
Distanciamientos (D)	2
Interacción C x D	6
(Tratamentos)	(11)
Bloques	4
Resíduo	44
Total	59

La separación de los grados de libertad queda así:

Fuentes de Variación	Grados de libertad
Entre cultivares dentro de D_1	3
Entre cultivares dentro de D_2	3
Entre cultivares dentro de D_3	3
Distanciamientos	2
(Tratamentos)	(11)
Bloques	4
Resíduo	44
Total	59

O alternativamente,

Fuentes de Variación	Grados de libertad
Entre distanciamientos dentro de C_1	2
Entre distanciamientos dentro de C_2	2
Entre distanciamientos dentro de C_3	2
Entre distanciamientos dentro de C_4	2
Cultivares	3
(Tratamentos)	(11)
Bloques	4
Resíduo	44
Total	59

8.1.3 Inconvenientes

a. El resultado del experimento y el análisis estadístico resultante son más complejos.

b. El número de tratamientos o combinaciones aumentan rápidamente, por ejemplo:

Factores	Niveles	Tratamientos
A y B	3 y 4	12
A y B	5 y 5	25
A, B y C	3, 4 y 2	24
A, B y C	3, 4 y 3	36

c. Con un gran número de combinaciones de tratamientos, tener unidades experimentales homogéneas es más difícil.

d. Convencidos de que algunas de las combinaciones de tratamientos pueden ser de muy poco o ningún interés, algunos de los recursos experimentales pueden ser malgastados.

Como alternativas para solucionar este problema, Barbin (2013) menciona las siguientes:

1. Usar factoriales fraccionados,
2. Usar bloques incompletos,
3. Usar confusión de los efectos, que consiste en repartir un bloque en 2 ó más sub-bloques, de modo que se confunda el efecto de alguna interacción sin interés práctico com el efecto de bloques.

Un ejemplo de interacción sin interés práctico es la interacción N × P × K (nitrógeno × fósforo × potasio) en ensayos de fertilizantes, que generalmente presenta F no significativa (Barbin, 2013).

Importante: en los ensayos sobre fertilizantes tienen mucha importancia los esquemas factoriales de las series 2^n y 3^n, especialmente los 2^3 y 3^3. La base indica el número de niveles del nutriente (o factor) y el exponente indica el número de nutrientes. Por tanto en el 2^3 tenemos los nutrientes: N, P y K, con dos niveles (0 y 1) cada uno. Las combinaciones del $2^3 = 2 \times 2 \times 2 = 8$.

8.1.4 Arreglos

Los tipos de arreglos más utilizados en experimentos factoriales son:

a) Combinatorio
b) Parcelas divididas
c) Parcelas subdivididas
d) Franjas divididas.

En el caso de experimentos trifactoriales, por ejemplo, se puede utilizar el arreglo en parcelas divididas. Considerando lo siguiente:

e) Ubicar un factor en la parcela grande, y los otros dos factores en arreglo combinatorio en la parcela pequeña.

f) Ubicar dos factores en arreglo combinatorio en la parcela grande, y el factor restante en la parcela pequeña.

8.2 EXPERIMENTOS FACTORIALES EN ARREGLO COMBINATORIO

8.2.1 Modelo estadístico

El modelo que se describe corresponde a un experimento bifactorial, en arreglo combinatorio dispuesto en un diseño en bloques completos al azar, debido a que es el más usado.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk} \quad \begin{cases} i = 1,2, \ldots, a \\ j = 1,2, \ldots, b \\ k = 1,2, \ldots, r \end{cases}$$

Siendo que:

Y_{ijk} = Variable de respuesta observada o medida en la ijk - ésima unidad experimental

μ = Media general

α_i = Efecto del i - ésimo nivel del factor "A"

β_j = Efecto del j - ésimo nivel del factor "B"

$(\alpha\beta)_{ij}$ = Efecto de la interacción entre el i - ésimo nivel del factor "A" y el j - ésimo nivel del factor "B"

γ_k = Efecto del k - ésimo bloque

ε_{ijk} = Error experimental asociado a la ijk - ésima unidad experimental

8.2.2 Hipótesis

Ho: $\alpha_i = 0$, para todo i, contra,
Ha: $\alpha_i \neq 0$, para algún i;

Ho: $\mu_{1..} = \ldots = \mu_{a..}$, contra
Ha: por lo menos $\mu_{i..} \neq \mu_{i'..}$ para $i \neq i'$;

Ho: $\beta_j = 0$, para todo j, contra,
Ha: $\beta_j \neq 0$, para algún j;

Ho: $\mu_{.1.} = \ldots = \mu_{.b.}$, contra
Ha: por lo menos $\mu_{.j.} \neq \mu_{.j'.}$ para $j \neq j'$;

Ho: $(\alpha\beta)_{ij} = 0$, para todo i y j, contra,
Ha: $(\alpha\beta)_{ij} \neq 0$, para algún i y j.

Ho: $\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu = 0$, contra
Ha: $\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu \neq 0$ para algún i y j.

8.2.3 Análisis de Varianza

Tabla de Datos para un experimento factorial 3 × 2 en bloques al azar.

Bloques	A_1		A_2		A_3		
	B_1	B_2	B_1	B_2	B_1	B_2	$Y_{..k}$
I	Y_{111}	Y_{121}	Y_{211}	Y_{221}	Y_{311}	Y_{321}	$Y_{..1}$
II	Y_{112}	Y_{122}	Y_{212}	Y_{222}	Y_{312}	Y_{322}	$Y_{..2}$
III	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
R	Y_{11r}	Y_{12r}	Y_{21r}	Y_{22r}	Y_{31r}	Y_{32r}	$Y_{..r}$
$Y_{ij.}$	$Y_{11.}$	$Y_{12.}$	$Y_{21.}$	$Y_{22.}$	$Y_{31.}$	$Y_{32.}$	$Y_{...}$

Tabla adicional para factores

Factor "B"	Factor "A"			$Y_{.j.}$
	A_1	A_2	A_3	
B_1	$Y_{11.}$	$Y_{21.}$	$Y_{31.}$	$Y_{.1.}$
B_2	$Y_{12.}$	$Y_{22.}$	$Y_{32.}$	$Y_{.2.}$
$Y_{i..}$	$Y_{1..}$	$Y_{2..}$	$Y_{3..}$	$Y_{...}$

$Y_{..k}$ = Total de bloques

$Y_{ij.}$ = Total de las interacciones

$Y_{i..}$ = Total del factor "A"

$Y_{.j.}$ = Total del factor "B"

$Y_{...}$ = Gran total

Cuadro resumen del Análisis de Varianza

FV	GL	SC	CM	Valor de F
Bloque	r – 1	$\sum_{k=1}^{r} \frac{Y_{..k}^2}{ab} - \frac{Y_{...}^2}{abr}$		
A	a – 1	$\sum_{i=1}^{a} \frac{Y_{i..}^2}{rb} - \frac{Y_{...}^2}{abr}$	SC_A/gl_A	CM_A/CM_{ee}
B	b – 1	$\sum_{j=1}^{b} \frac{Y_{.j.}^2}{ra} - \frac{Y_{...}^2}{abr}$	SC_B/gl_B	CM_B/CM_{ee}

AB	(a – 1) (b– 1)	$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{r}-\frac{Y_{...}^2}{abr}$ $–SC_A – SC_B$	SC_{AB}/gl_{AB}	CM_{AB}/CM_{ee}
Error	ab – 1 (r – 1)	$SC_{TOTAL}– SC_{BLOQUES} – SC_A– SC_B–SC_{AB}$	Sc_{ee}/gl_{ee}	
Total	abr – 1	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}Y_{ijk}^2-\frac{Y_{...}^2}{abr}$		

Los valores de F que aparecen en la tabla sirven para probar hipótesis de la siguiente manera:

1. Para el factor A

   $F = CM_A/CM_{ee}$; $F_{crítica\ (a-1;\ glee;\ \alpha)}$

2. Para el factor B

   $F = CM_B/CM_{ee}$; $F_{crítica\ (b-1;\ glee;\ \alpha)}$

3. Para la interacción AB

   $F = CM_{AB}/CM_{ee}$; $F_{crítica\ (a-1,\ b-1;\ glee;\ \alpha)}$

Regla de Decisión

Si $F \geq F_{crítica}$ Se rechaza la hipótesis nula
Si $F < F_{crítica}$ No se rechaza la hipótesis nula

8.2.4 Análisis post-ANOVA

Para efectuar la prueba de Tukey o SNK seguir el procedimiento anteriormente descrito, considerando los comparadores de la manera siguiente:

1. Para el factor A

$Wp = q_{(a,glee,\alpha)} \times S_{\bar{x}}$, siendo

$$S_{\bar{x}} = \sqrt{\frac{CMee}{rb}}$$

q = valor obtenido en la tabla de Tukey
a = niveles del factor A.
α = nivel de significancia
CMee = cuadrado medio del error
r = número de repeticiones
b = número de niveles del factor B.

2. Para el factor B

$$Wp = q_{(b,glee,\alpha)} \times S_{\bar{x}} \text{, siendo: } S_{\bar{x}} = \sqrt{\frac{CMee}{ra}}$$

3. Para la interacción AB

$$Wp = q_{(ab,glee,\alpha)} \times S_{\bar{x}} \text{, siendo: } S_{\bar{x}} = \sqrt{\frac{CMee}{r}}$$

¿A qué factor o interacción debe aplicarle la prueba de Tukey? Consulte la siguiente tabla:

Efecto	Resultado de la Prueba							
A	*	NS	*	NS	NS	*	*	NS
B	*	*	NS	NS	*	NS	*	NS
AB	*	*	*	*	NS	NS	NS	NS
Efectuar la prueba a:	AB	AB	AB	AB	B	A	A ; B	----

***** : “diferencias significativas”
NS : “no se presentaron diferencias significativas”

8.2.5 Ejemplo de Aplicación 1

Se estudió el efecto de diferentes dosis de fertilizante fosforado sobre dos tipos de planta de haba (*Vicia faba)*. Se pensó que los tipos de planta bien podían responder en forma diferente a la fertilización, así que se decidió llevar a cabo un experimento factorial con dos factores:

Factor	Nivel		
	1	2	3
A. Tipo de Planta	Corto	Alto	
B. Dósis de Fósforo	0 kg/ha P_2O_5	25 kg/ha P_2O_5	50 kg/ha P_2O_5

El experimento se llevó a cabo en un diseño en bloques completos al azar en arreglo combinatorio, con cuatro repeticiones. Los resultados (producción en kg/ha) se aprecian en el cuadro 1.

Cuadro 1: Producción (kg/ha) de dos tipos de haba bajo diferentes niveles de fertilización de fósforo.

Tipo de Planta	Dosis de Fósforo	Bloques				$Y_{ij.}$
		I	II	III	IV	
Corto	0	11.5	13.6	14.3	14.5	53.90
	25	17.1	17.6	17.6	18.1	70.40
	50	18.2	17.6	18.2	18.9	72.90

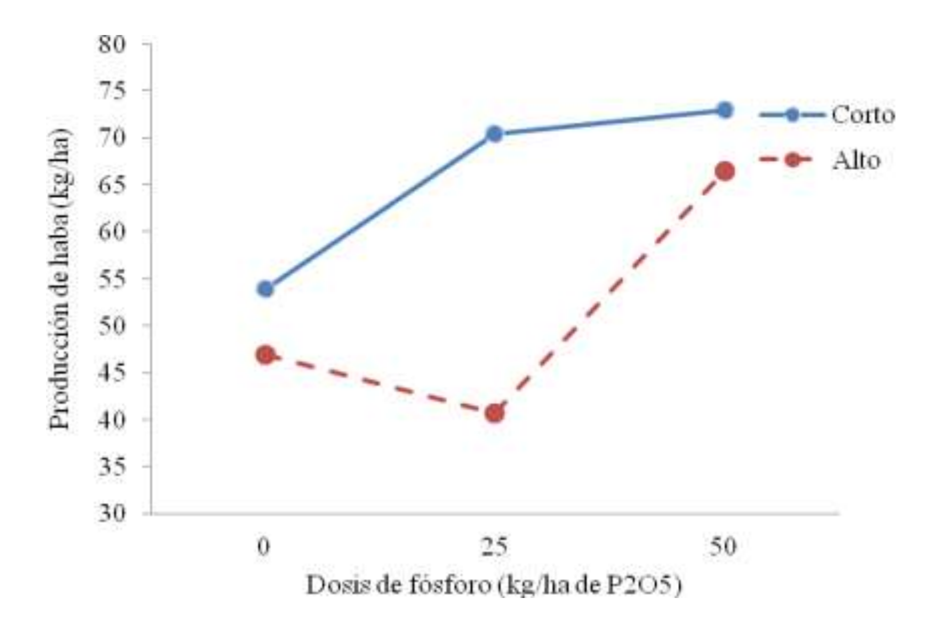
Largo	0	11.0	11.2	12.1	12.6	46.90
	25	8.3	10.5	9.1	12.8	40.70
	50	15.7	16.7	16.6	17.5	66.50
	$Y_{..k}$	81.8	87.2	87.9	94.4	

Solución: $Y_{...} = 351.30$ $\bar{Y}_{...} = 14.64$

Tabla adicional:

Dosis de fósforo	Tipo de planta		$Y_{.j.}$	$\bar{Y}_{.j.}$
	A_1	A_2		
B_1	53.90	46.90	100.80	12.60
B_2	70.40	40.70	111.10	13.89
B_3	72.90	66.50	139.40	17.43
$Y_{i..}$	197.20	154.10		
$\bar{Y}_{i..}$	16.43	12.84		

El gráfico para la interacción se presenta a continuación:



Note la presencia de un aumento de la producción a medida que aumenta la dosis de fósforo y que el tipo de planta corto supera al tipo de planta alto.

$$\overline{Y}_{i..} = \frac{\sum_{i=1}^{a} Y_{i..}}{rb} \qquad \overline{Y}_{.j.} = \frac{\sum_{j=1}^{b} Y_{.j.}}{ra} \qquad \overline{Y}_{...} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}}{r}$$

$$SC_A = \frac{197.2^2 + 154.10^2}{(4)(3)} - \frac{351.30^2}{(4)(3)(2)} = 77.40$$

$$SC_B = \frac{100.8^2 + 111.10^2 + 139.40^2}{(4)(2)} - \frac{351.30^2}{(4)(3)(2)} = 99.87$$

$$SC_{AB} = \frac{53.9^2 + 70.40^2 + \ldots + 66.50^2}{4} - \frac{351.30^2}{(4)(3)(2)} - 77.40 - 99.87 = 44.10$$

$$SC_{BLOQUES} = \frac{81.80^2 + 87.20^2 + 87.90^2 + 94.40^2}{(3)(2)} - \frac{351.30^2}{(4)(3)(2)} = 13.32$$

$$SC_{TOTAL} = 11.5^2 + 17.10^2 + \ldots + 17.50^2 - \frac{351.30^2}{(4)(3)(2)} = 243.38$$

$$SC_{EE} = 243.38 - 77.40 - 99.87 - 44.1 - 13.32 = 8.68$$

Cuadro resumen del Análisis de Varianza

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
Bloque	3	13.32			
A	1	77.40	77.40	133.45*	4.54
B	2	99.87	49.94	86.10*	3.68
AB	2	44.11	22.06	38.03*	3.68
Error Experimental	15	8.68	0.58		
Total	23	243.38			

$$CV = \frac{\sqrt{0.58}}{14.64} \times 100 = 5.20\%$$

Prueba de comparación múltiple de medias de acuerdo con el criterio de Tukey, para la interacción

a) Medias de las interacciones

Interacción	$\overline{Y}_{ij.}$
A_1B_1	13.48
A_1B_2	17.60
A_1B_3	18.23
A_2B_1	11.73
A_2B_2	10.18
A_2B_3	16.63

b) Comparador

$$Wp = q_{(6,15,0.05)} \times \sqrt{\frac{0.58}{4}} \quad = \quad 4.59 \times 0.3807 \; = \quad 1.7474$$

c) Matriz de diferencias

Interacción	Media	A_1B_3 18.23	A_1B_2 17.6	A_2B_3 16.63	A_1B_1 13.48	A_2B_1 11.73	A_2B_2 10.18
A_2B_2	10.18	8.05*	7.42*	6.45*	3.30*	1.55	—
A_2B_1	11.73	6.50*	5.87*	4.9*	1.75*	—	
A_1B_1	13.48	4.75*	4.12*	3.15*	—		
A_2B_3	16.63	1.60	0.97	—			
A_1B_2	17.6	0.63	—				
A_1B_3	18.23	—					

d) Presentación de resultados

Interacción	Producción promedio (kg/ha) de haba	Grupo Tukey
(A_1B_3) Corto × 50 kg/ha	18.23	A
(A_1B_2) Corto × 25 kg/ha	17.60	A
(A_2B_3) Alto × 50 kg/ha	16.63	A
(A_1B_1) Corto × 0 kg/ha	13.48	B
(A_2B_1) Alto × 0 kg/ha	11.73	c
(A_2B_2) Alto × 25 kg/ha	10.18	c

8.2.6 Ejemplo de Aplicación 2

Vamos a considerar los datos de un experimento realizado en un diseño completamente al azar con 4 repeticiones, en arreglo combinatorio 3 × 2, para evaluar los efectos de 3 recipientes (R_1, R_2 y R_3) para producción de plántulas y 2 especies de eucaliptos (E_1 y E_2), la variable de respuesta medida fue la altura media de las plántulas, en cm, a los 80 días de edad. Los recipientes y las especies evaluadas fueron:

R_1 = bolsa plástica pequeña
R_2 = bolsa plástica grande, y
R_3 = laminado.

E_1 = *Eucalyptus citriodora* y
E_2 = *Eucalyptus grandis*.

Los resultados se presentan en el cuadro siguiente:

Especies	Recipientes	Repeticiones				$Y_{ij.}$
		I	II	III	IV	
E_1	R_1	26.2	26.0	25.0	25.4	102.6
	R_2	25.7	26.3	25.1	26.4	103.5
	R_3	22.8	19.4	18.8	19.2	80.2
E_2	R_1	24.8	24.6	26.7	25.2	101.3
	R_2	19.6	21.1	19.0	18.6	78.3
	R_3	19.8	21.4	22.8	21.3	85.3

Estos datos fueron reportados por Banzatto y Kronka (2011) del trabajo "Métodos de producción de plántulas de eucalipto", realizado por Simões (1970). Los cálculos para el análisis de varianza se presentan a continuación:

$$SC_{TOTAL} = 26.2^2 + 26.0^2 + \ldots + 21.3^2 - \frac{551.20^2}{2 \times 3 \times 4} = 198.79$$

$$SC_{TRATS} = \frac{\sum_{i=1}^{t} T_i^2}{r} = \frac{102.6^2 + 103.5^2 + \ldots + 85.3^2}{4} - \frac{551.20^2}{2 \times 3 \times 4} = 175.70$$

El análisis de varianza preliminar de los datos de altura de las plántulas a los 80 días es presentado a continuación:

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
Tratamientos	5	175.70	35.14	27.45**	2.77
Resíduo	18	23.09	1.28		
Total	23				

Verificamos que existen diferencias significativas, indicando que los tratamientos presentan diferentes efectos sobre las alturas de las plántulas. Debemos proceder al desdoblamiento de los 5 grados de libertad de los tratamientos, para estudiar los efectos de los recipientes (R), de las especies (E) y de la interacción R × E, se la siguiente forma:

Tratamientos…. 5 gl	Recipientes (R)	2 gl
	Especies (E)	1 gl
	Interacción R × E	2 gl

Para el cálculo de las sumas de cuadrados correspondientes a los efectos principales de los factores y a la interacción entre ellos, debemos organizar un cuadro auxiliar, que relaciona los niveles de los dos factores:

	R_1	R_2	R_3	Total de E
E_1	102.6	103.5	80.2	286.3
E_2	101.3	78.3	85.3	264.9
Total de R	203.9	181.8	165.5	551.2

$$SC_R = \frac{203.9^2 + 181.8^2 + 165.5^2}{2 \times 4} - \frac{551.20^2}{2 \times 3 \times 4} = 92.86$$

$$SC_E = \frac{286.3^2 + 264.9^2}{3 \times 4} - \frac{551.20^2}{2 \times 3 \times 4} = 19.08$$

$$SC_{R \times E} = SC_{TRATS} - SC_R - SC_E = 175.70 - 92.86 - 19.08 = 63.76$$

El análisis de varianza con desboblamiento de los grados de libertad de los tratamientos, de acuerdo con el arreglo combinatorio 3 × 2, es presentado en el cuadro siguiente:

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
Recipientes (R)	2	92.86	46.43	36.27**	3.55
Especies (E)	1	19.08	19.08	14.91**	4.41
Interacción R × E	2	63.76	31.88	24.91**	3.55
Tratamientos	5	(175.70)			
Resíduo	18	23.09	1.28		
Total	23				

Verificamos por la prueba de F, que la Interacción R × E fue significativa ($p<0.05$), indicando que existe una dependencia entre los efectos de los factores Recipientes (R) y Especies (E).

Entonces las conclusiones que podemos obtener para los efectos principales de Recipientes (R) y de Especies (E) quedan perjudicadas, pues:

- Los efectos de los recipientes dependen de la especie utilizada; o
- Los efectos de las especies dependen del recipiente utilizado.

Entonces, debemos proceder al desdoblamiento de la interacción R × E, lo que puede ser realizado de dos maneras: a) para estudiar el comportamiento de las especies dentro de cada recipiente; b) para estudiar el comportamiento de los recipientes dentro de cada especie.

a) Desdoblamiento de la interacción R × E para estudiar el comportamiento de las Especies dentro de cada Recipiente

$$\text{SC. Especies d. } R_1 = \frac{102.6^2 + 101.3^2}{4} - \frac{203.9^2}{2 \times 4} = 0.21$$

$$\text{SC. Especies d. } R_2 = \frac{103.6^2 + 78.3^2}{4} - \frac{181.8^2}{2 \times 4} = 79.38$$

$$\text{SC. Especies d. } R_3 = \frac{80.2^2 + 85.3^2}{4} - \frac{165.5^2}{2 \times 4} = 3.25$$

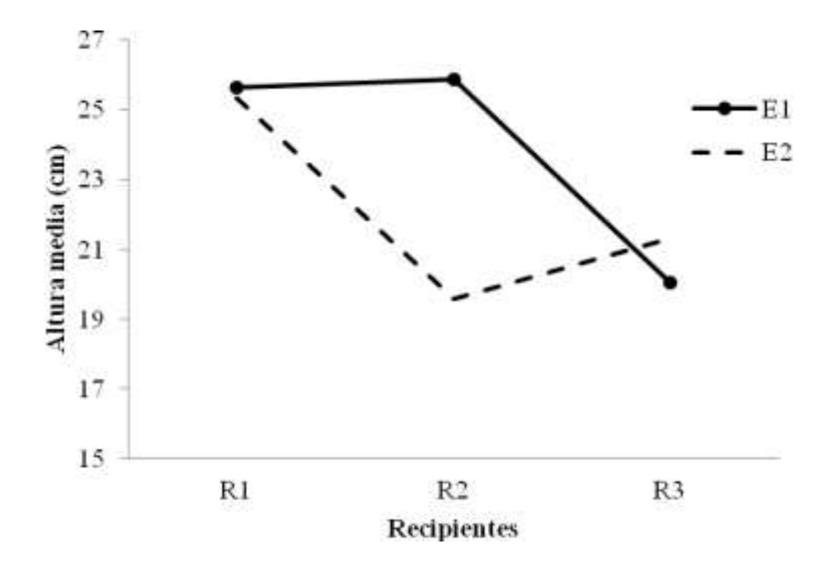
Se puede verificar que: SC Especies dentro de R_1 + SC Especies dentro de R_2 + SC Especies dentro de R_3 = SC Especies + SC R × E = 0.21+79.38+3.25 = 19.08 + 63.76 = 82.84. El cuadro de ANOVA se presenta a continuación:

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
SC Especies d. R_1	1	0.21	0.21	0.16	4.41
SC Especies d. R_2	1	79.38	79.38	62.02**	4.41
SC Especies d. R_3	1	3.25	3.25	2.54	4.41
Residuo	18	23.09	1.28		

Conclusiones:

a) Cuando se utiliza el Recipiente 1 (bolsa plástica pequeña), no hay diferencia significativa en el crecimiento de las plántulas de las 2 especies.
b) Cuando se utiliza el Recipiente 2 (bolsa plástica grande), hay diferencia significativa en el crecimiento de las plántulas de las 2 especies, siendo mejor para la especie E_1 (*E. citriodora*)
c) Cuando se utiliza el Recipiente 3 (laminado), no hay diferencia significativa en el crecimiento de las plántulas de las 2 especies.

Gráficamente:



b) Desdoblamiento de la interacción R × E para estudiar el comportamiento de los Recipientes dentro de cada Especie

$$SC.\text{Recipientes d. } E_1 = \frac{102.6^2 + 103.5^2 + 80.2^2}{4} - \frac{286.3^2}{3 \times 4} = 87.12$$

$$SC.\text{Recipientes d. } E_2 = \frac{101.3^2 + 78.3^2 + 85.3^2}{4} - \frac{264.9^2}{3 \times 4} = 69.50$$

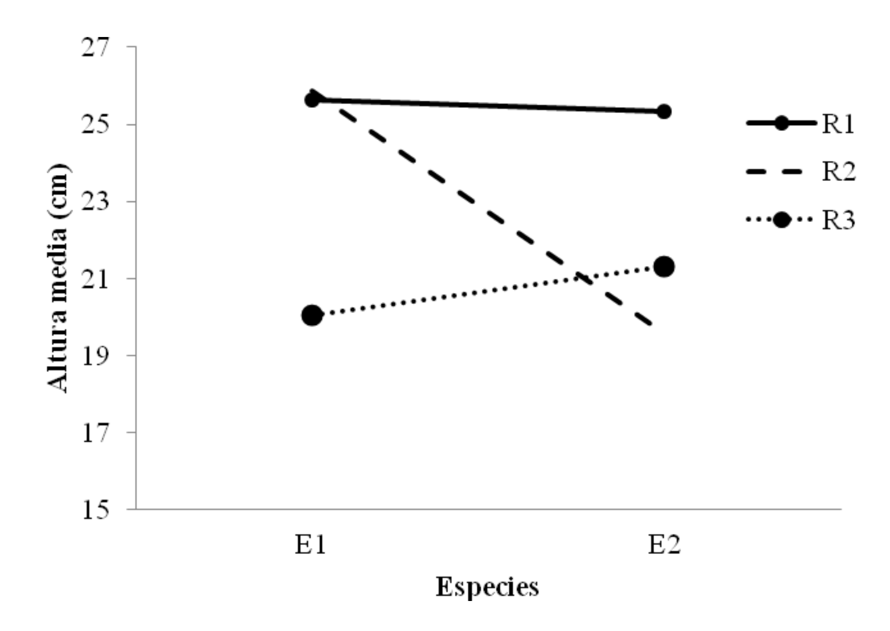
Se puede verificar que: SC Recipientes dentro de E_1 + SC Recipientes dentro de E_2 = SC Recipientes + SC R × E = 87.12+69.50 = 92.86 + 63.76 = 156.62. El cuadro de ANOVA se presenta a continuación:

Fuentes de Variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Valor de F	F crítica
SC Recipientes d. E_1	2	87.12	43.56	34.03**	3.55
SC Recipientes d. E_2	2	69.50	34.50	27.15**	3.55
Residuo	18	23.09	1.28		

Conclusiones:

a) Los 3 recipientes tienen efectos diferentes sobre el crecimiento de las plántulas de la Especie 1 (*E. citriodora*)
b) Los 3 recipientes tienen efectos diferentes sobre el crecimiento de las plántulas de la Especie 2 (*E. grandis*)

Gráficamente:



Debemos entonces, comparar las medias de los recipientes:

a) Dentro de la Especie 1 (*E. citriodora*)
b) Dentro de la Especie 2 (*E. grandis*)

Vamos a calcular esas medias y comparárlas por medio de la prueba de Tukey

a) Recipientes dentro de Especie 1 (*E. citriodora*)

$$\hat{m}_{R_1E_1} = \frac{102.6}{4} = 25.7 \text{ cm} \qquad \hat{m}_{R_2E_1} = \frac{103.5}{4} = 25.9 \text{ cm} \qquad \hat{m}_{R_3E_1} = \frac{80.2}{4} = 20.1 \text{ cm}$$

$$\hat{s}(\hat{m}) = \frac{s}{\sqrt{r}} = \frac{\sqrt{1.28}}{\sqrt{4}} = 0.6 \text{ cm} \qquad \Delta = q\frac{s}{\sqrt{r}} = 3.65 \times 0.6 = 2.2 \text{ cm}$$

El valor q se obtuvo considerando 3 niveles de R, 18 grados de libertad del residuo y 5% de significancia.

	$\hat{m}_{R_2E_1}$	$\hat{m}_{R_1E_1}$	$\hat{m}_{R_3E_1}$
$\hat{m}_{R_2E_1}$ **a**	-	0.2 ^{NS}	5.8 *
$\hat{m}_{R_1E_1}$ **a**	-	-	5.6 *
$\hat{m}_{R_3E_1}$ **b**	-	-	-

Para la Especie 1 (*E. citriodora*), los mejores recipientes fueron: R_1 (bolsa plástica pequeña) y R_2 (bolsa plástica grande), que determinaron el crecimiento de las plántulas significativamente mayores que R_3 (laminado), sin diferir entre si.

b) Recipientes dentro de Especie 2 (*E. grandis*)

$$\hat{m}_{R_1E_2} = \frac{101.3}{4} = 25.3\,\text{cm} \qquad \hat{m}_{R_2E_2} = \frac{78.3}{4} = 19.6\,\text{cm} \qquad \hat{m}_{R_3E_2} = \frac{85.3}{4} = 21.3\,\text{cm}$$

$$\hat{s}(\hat{m}) = 0.6\,\text{cm} \qquad \Delta = 2.2\,\text{cm}$$

	$\hat{m}_{R_1E_2}$	$\hat{m}_{R_3E_2}$	$\hat{m}_{R_2E_2}$
$\hat{m}_{R_1E_2}$ **a**	-	4.0^{*}	5.7 *
$\hat{m}_{R_3E_2}$ **b**	-	-	1.7^{NS}
$\hat{m}_{R_2E_2}$ **b**	-	-	-

Para la Especie 2 (*E. grandis*), el mejor recipiente fue: R_1 (bolsa plástica pequeña) que determino el crecimiento de las plántulas significativamente mayores que R_2 (bolsa plástica grande) y R_3 (laminado). Las medias de R_2 y R_3 (dentro de Especie 2) no difieren significativamente entre si.

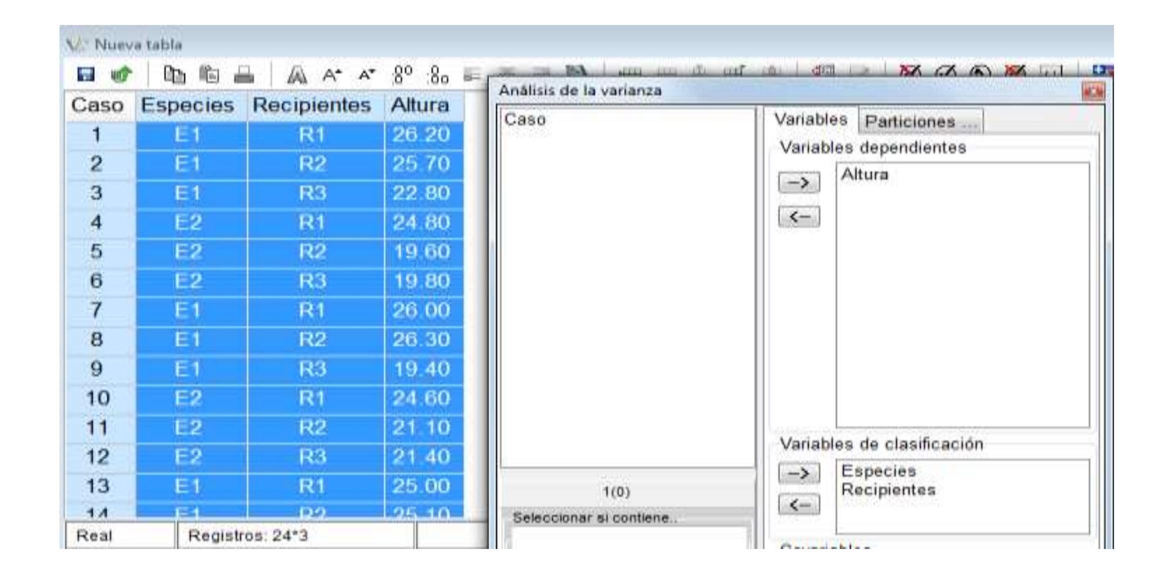
Los resultados del experimento pueden resumirse en el cuadro siguiente:

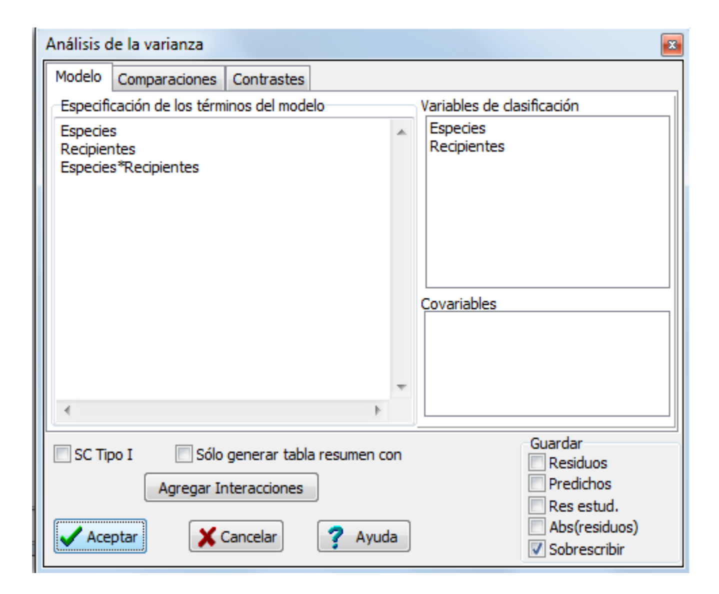
	R_1	R_2	R_3
E_1	25.7 **a A**	25.9 **a A**	20.1 **b A**
E_2	25.3 **a A**	19.6 **b B**	21.3 **b A**

a, b Para cada Especie, medias de Recipientes seguidas de la misma letra minúscula no difieren significativamente entre si.

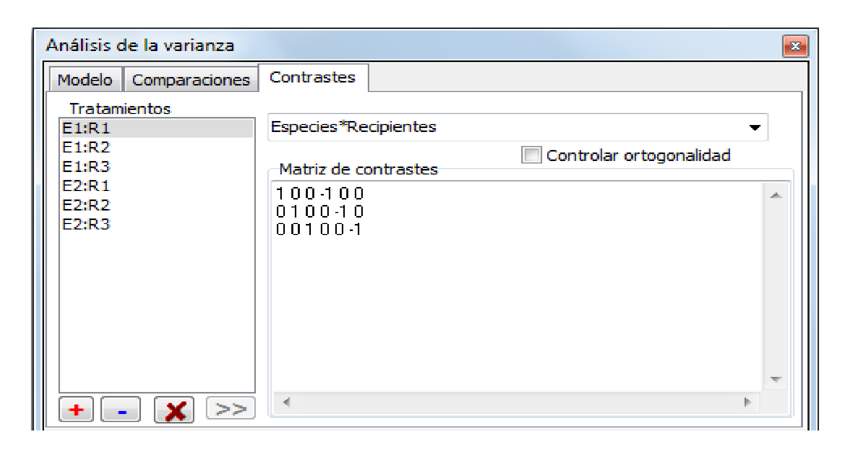
A, B Para cada Recipiente, medias de Especies seguidas de la misma letra mayúscula no difieren entre significativamente entre si.

Para resolver el ejercicio anterior en Infostat v.2015, siga lo mostrado en las ventanas siguientes:





a) Desdoblamiento de la interacción R × E para estudiar el comportamiento de las Especies dentro de cada Recipiente



Contrastes

Especies*Recipientes	Contraste	E.E.	SC	gl	CM	F	p-valor
Contraste1	0.32	0.80	0.21	1	0.21	0.16	0.6897
Contraste2	6.30	0.80	79.38	1	79.38	61.88	<0.0001
Contraste3	-1.28	0.80	3.25	1	3.25	2.53	0.1288
Total			82.84	3	27.61	21.53	<0.0001

Coeficientes de los contrastes

Especies*Recipientes	Ct.1	Ct.2	Ct.3
E1:R1	1.00	0.00	0.00
E1:R2	0.00	1.00	0.00
E1:R3	0.00	0.00	1.00
E2:R1	-1.00	0.00	0.00
E2:R2	0.00	-1.00	0.00
E2:R3	0.00	0.00	-1.00

La salida generada por Infostat v. 2015 se presenta a continuación:

Análisis de la varianza

Variable	N	R²	R² Aj	CV
altura	24	0.88	0.85	4.93

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	175.70	5	35.14	27.39	<0.0001
Especie	19.08	1	19.08	14.88	0.0012
Recip	92.86	2	46.43	36.20	<0.0001
Especie*Recip	63.76	2	31.88	24.85	<0.0001
Error	23.09	18	1.28		
Total	198.79	23			

8.2.7 Programa en SAS para un experimento bifactorial con arreglo combinatorio

```
OPTIONS nodate nonumber; /*Para que en la salida no aparezca fecha ni paginación*/
DATA combin;
INPUT bloque planta $ fosforo rend;
CARDS;
1   corto   0    11.5
1   corto   25   17.1
1   corto   50   18.2
1   largo   0    11
1   largo   25   8.3
1   largo   50   15.7
2   corto   0    13.6
2   corto   25   17.6
2   corto   50   17.6
2   largo   0    11.5
2   largo   25   10.5
2   largo   50   16.7
3   corto   0    14.3
3   corto   25   17.6
3   corto   50   18.2
3   largo   0    12.1
3   largo   25   9.1
3   largo   50   16.6
4   corto   0    14.5
4   corto   25   18.1
4   corto   50   18.9
4   largo   0    12.6
4   largo   25   12.8
4   largo   50   17.5
;
PROC GLM;
   CLASS bloque planta fosforo;
   MODEL rend=bloque planta fosforo planta*fosforo/SS1;
   LSMEANS planta*fosforo/PDIFF;
RUN;
****************************
```

Dado que en este análisis la interacción entre tipo de planta y dosis de fósforo es significativa, podemos estudiar si hay diferencia entre las dosis de fósforo para cada tipo de planta (con Tukey):

```
PROC GLM;
   CLASS bloque planta fosforo;
   MODEL rend=bloque planta fosforo planta*fosforo/SS1;
   LSMEANS planta*fosforo/slice=planta adjust=tukey PDIFF=all;
RUN;
```

8.2.7 Ejercicios propuestos

1. Se realizó un ensayo en el departamento de Chimaltenango en el cual se evalúo la respuesta del rendimiento del cultivo de haba *(Vicia faba)* frente a cuatro niveles de Nitrógeno y cuatro niveles de Fósforo, en el área existía una gradiente de variabilidad por lo cual se utilizó el diseño de bloques al azar. Los resultados se presentan en los siguientes cuadros:

Niveles de los factores evaluados

FACTOR	NIVEL (kg/ha)	CODIGO
Nitrógeno	0	0
	50	1
	100	2
	150	3
Fósforo	0	0
	50	1
	100	2
	150	3

Rendimiento en (kg/ha) de haba (Vicia faba)

Niveles de Nitrógeno	Niveles de fósforo	Repeticiones			
		I	II	III	IV
0	0	833	450	450	1083
1	0	450	834	450	833
2	0	450	833	667	833
3	0	1083	833	917	833
0	1	1250	1083	917	1250
1	1	1083	1083	583	1083
2	1	1083	1583	1167	1000
3	1	917	1167	1000	1250
0	2	833	583	750	1333
1	2	1167	833	1500	750
2	2	1417	1167	1000	833
3	2	1583	1500	583	1417
0	3	750	450	667	417
1	3	1500	1167	667	1083
2	3	1333	917	833	1167
3	3	1167	833	1083	1250

a) Construya una gráfica para mostrar la posible interacción entre los niveles de los factores nitrógeno y fósforo.

b) Realice el ANOVA y concluya

c) En caso de ser necesario realice el análisis post-ANOVA

2. Realize el análisis de varianza para el siguiente experimento bifactorial en Bambú (*Bambusa arundinacea*) en el que se consideraron dos espaciamientos (Factor A) y 3 edades de los rizomas para la plantación (Factor B), dispuesto en un diseño de bloques completos al azar con tres repeticiones (Jayaraman, K. 1999). Los niveles de cada factor y sus combinaciones, se presentan en el cuadro siguiente.

Edad de los rizomas para la plantación (meses)	Espaciamiento (metros)	
	10 m x 10 m (A_1)	12 m x 12 m (A_2)
6 (B_1)	$A_1 B_1$	$A_2 B_1$
12 (B_2)	$A_1 B_2$	$A_2 B_2$
24 (B_3)	$A_1 B_3$	$A_2 B_3$

La variable de respuesta medida fue altura máxima promedio de los tallos (cañas), expresada en centímetros. Los resultados se presentan a continuación:

Combinación	Repeticiones		
	I	II	III
$A_1 B_1$	46.50	55.90	78.70
$A_1 B_2$	49.50	59.50	78.70
$A_1 B_3$	127.70	134.10	137.10
A_2B_1	49.30	53.20	65.30
$A_2 B_2$	65.50	65.00	74.00
$A_2 B_3$	67.90	112.70	129.00
Total Y..k			

a) Construya una gráfica para mostrar la posible interacción entre los niveles de los factores edad de los rizomas y espaciamiento.
b) Realice el ANOVA y concluya
c) En caso de ser necesario realice el análisis post-ANOVA

3. Rivera, F. (2001) realizó el trabajo de tesis de grado "Evaluación de N y P, en el rendimiento de cardamomo *(Elettaria cardamomun* M.) en la serie de suelos Tamahú, aldea Choval, Cobán, Alta Verapaz". Los niveles de nitrógeno evaluados fueron: 0, 50, 100 y 150 kg/ha/año; y los de fósforo: 0, 40, 80 y 120 kg/ha/año. La variable de respuesta medida fue la producción de cardamomo en cereza, expresada en kg/planta. El resumen del ANDEVA se presenta a continuación:

FV	**GL**	**SC**	**CM**	**Valor de F**	**Pr > F**
Repetición	2	0.20820634			
Nitrógeno	3	24.81599830	8.27199943	36.89	0.0001
Fósforo	3	6.66365625	2.22121975	9.91	0.0001
N x P	9	12.50315497	1.38923944	6.20	0.0001
Residuo	30	6.72683850	0.22422795		
Total	47	50.91785737			

Las medias para las interacciones se presentan en el cuadro siguiente:

Nitrógeno	Fósforo	Rendimiento Promedio
0	0	1.89942333
0	40	2.42085333
0	80	2.45696667
0	120	2.02227000
50	0	2.47586667
50	40	2.71211333
50	80	2.27741667
50	120	2.81606000
100	0	3.05296000
100	40	4.55483333
100	80	4.16739000
100	120	4.06344333
150	0	2.50773000
150	40	4.73571667
150	80	4.62098333
150	120	2.49463333

a) Interprete el ANOVA y concluya
b) Analice el comportamiento de los niveles de fósforo en presencia de cada uno de los niveles de nitrógeno (sugerencia: ajuste un modelo de regresión)

4. Alfaro, M. R. (1999) realizó el trabajo de tesis "Evaluación inicial del efecto de tres intensidades de raleo y tres de poda en el crecimiento de una plantación de *Pinus caribaea Morelet* Var. *hondurensis*, en Livingston, Izabal".

En esta investigación se utilizó un diseño de bloques completos al azar y se evaluaron dos factores: Factor A (raleos): se utilizó el índice Hart (S%), con intensidades: S%= actual, S%= 22 y S%=28, debido a que el índice de Hart establece que entre 20 y 30% del espaciamiento, una plantación puede crecer adecuadamente sin competencia. Factor B (poda): se hizo con las siguientes intensidades: sin poda (SP), ⅓ y ½ de la altura de la copa viva. Las variables de respuesta medidas fueron: diámetro, altura, área basal y volumen. A continuación se presentan los datos para la variable volumen (m^3), correspondientes a la tercera medición.

Poda	Raleos	Repeticiones			
		I	II	III	IV
S.P.	S%A	0.0803	0.0569	0.0548	0.0691
	S%22	0.0965	0.0643	0.0673	0.0589
	S%28	0.1130	0.0833	0.0612	0.0581
⅓h	S%A	0.1020	0.0542	0.0594	0.0466
	S%22	0.0860	0.0756	0.0869	0.0488
	S%28	0.1390	0.0857	0.0793	0.0872
½h	S%A	0.0739	0.0625	0.0577	0.0417
	S%22	0.0953	0.0635	0.0105	0.0708
	S%28	0.1300	0.0558	0.0811	0.1160

a) Realice el ANOVA y concluya
b) En caso de ser necesario realice el análisis post-ANOVA

5. Un Ingeniero Forestal realizó un experimento con el objetivo de estudiar la influencia del tipo de aparato y del operador del mismo, en la medición de alturas (metros) de árboles de *Eucaliptus saligna* de 7 años de edad. Fueron evaluados 5 diferentes aparatos, por 4 operadores (factorial 5 × 4), en 10 árboles (cada árbol fue considerado como un bloque). Los resultados se presentan a continuación:

		Árbol									
Operario	Aparato	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1	22.4	20.85	23.6	21	19.1	19.8	16.55	14.75	21.1	14.3
	2	22.9	21.4	23.95	22.25	21.4	21	16.9	14.85	22	15
	3	23.5	21	23.75	20.75	19.5	19.5	17.5	14.5	20	14
	4	22.5	20.5	23.2	21	21	18.9	17.8	14.3	20.6	14.2
	5	21.45	19.2	23.35	20.35	19.95	19.35	17.45	14.45	22	14.75
2	1	22.65	20.65	23	20.75	20.25	19.8	17.25	15	19.75	14.25
	2	23	20.7	22.5	20.95	22.25	20.75	18	14.75	20.5	15.25
	3	22	19.5	23.25	20.5	21.25	19.75	17.75	14.75	20.5	14.25
	4	22.9	21.2	24.6	21.5	21.2	20	18.7	15	21.5	14.2
	5	21.45	18.9	23.2	20.25	19.95	19.2	17.35	14.35	21.8	14.65
3	1	22.5	21.25	23.1	20.6	21	19.5	16.6	14.35	20.75	14.1
	2	22.5	21	23	21.75	22.75	20.35	17.2	14.85	22.35	16
	3	22.75	20.5	22.75	19.5	20.5	19.75	17.25	14.25	21.5	14.25
	4	21.75	19.35	21.75	19.5	20.5	19	16.35	14.1	20.85	13.85
	5	21.35	19.2	23.2	20.3	20	19.3	17.5	14.4	21.9	14.8
4	1	21.25	21.25	22.25	21.25	18	20	17.25	14.65	21	14.25
	2	22.1	21.6	22.35	21.75	19.75	20.65	16.7	15.75	20.85	15.4
	3	21.25	21.5	22.1	21.7	19.75	19.75	18.2	14.6	21.25	14.75
	4	21.9	21	22.75	20.75	19.7	20	18.45	14.3	20.75	15.1
	5	21.2	18.9	23.3	20.3	19.9	19.3	17.4	14.5	22	14.4

Fuente: Barbin (2013)

Aparatos: 1. Hipsómetro de Blume-Leis, 2. Hipsómetro de Haga, 3. Hipsómetro Weise, 4. Plancheta dasométrica y 5. Trena.

a) Realice el ANOVA, indicando las hipótesis evaluadas, y concluya
b) En caso de ser necesario realice el análisis post-ANOVA.

6. Los datos que se presentan a continuación se refieren al diámetro medio (cm) a los 5 años de edad, de las plantas útiles de cada parcela, de un ensayo factorial 2 × 4, en bloques completos al azar con 3 repeticiones, realizado entre 1966 y 1971 por H.A. Mello en el Huerto Santa Terezinha en Mogi-Iguaçu, un municipio brasileño del Estado de Sao Paulo que está a una altitud de 591 metros. Cada parcela era constituida de 224 árboles (14×16) para el distanciamiento 3.0 × l.5 m y de 168 árboles (14×12) para el de 3.0 × 2.0 m. Solamente 168 e 120 árboles, respectivamente, fueron usados, dejando una bordadura simple en cada parcela. Las especies de Eucalipto evaluadas fueron: *E. saligna*, *E. grandis*, *E. alba* y *E. propinqua*.

Distanciamiento	*E. saligna*	*E. grandis*	*E. alba*	*E. propinqua*
3.0 × 1.5 m	10.69	9.68	9.59	9.40
	10.05	10.58	9.78	7.50
	10.31	10.44	10.17	8.66
3.0 × 2.0 m	12.95	11.78	11.46	10.02
	10.89	12.28	10.89	10.45
	12.37	12.82	10.97	10.71

Además foi medido o rendimiento de madera seca a los 5 años de edad, expresados en toneladas por hectárea:

Distanciamiento	*E. saligna*	*E. grandis*	*E. alba*	*E. propinqua*
3.0 × 1.5 m	79.65	83.97	83.97	55.86
	88.94	84.93	84.93	55.16
	80.54	87.53	87.53	59.05
3.0 × 2.0 m	89.21	81.23	72.98	59.71
	94.40	93.51	92.11	61.41
	95.81	81.13	83.62	53.43

Y la altura media (metros) de 10 plantas de 5 años de edad, por cada parcela.

Distanciamiento	*E. saligna*	*E. grandis*	*E. alba*	*E. propinqua*
3.0 × 1.5 m	15.25	15.45	15.70	14.90
	17.14	12.30	17.40	13.90
	16.00	17.85	16.95	14.45
3.0 × 2.0 m	19.20	15.55	16.90	14.90
	15.80	17.45	17.95	15.80
	17.05	16.90	17.80	16.65

Fuente: Mello, H.A.; Simões, J.W.; Mascarenhas, J.; Couto, H. T. 1971. Influência do espaçamento na produção de madeira de eucalipto em solo de cerrado. IPEF n.2/3, p.3-30. Disponible en: http://www.ipef.br/publicacoes/scientia/nr02-03/cap01.pdf

Para cada variable realice lo siguiente:

a) Verifique gráficamente si existe interacción entre especies x distanciamiento.
b) Haga el Análisis de Varianza.
c) Haga el desdoblamiento del número de grados de libertad de especies (S) más el de la interacción (S × E). Obtenga conclusiones.
d) Haga el desdoblamiento del número de grados de libertad de distanciamientos (E) más el de la interacción (S × E). Obtenga conclusiones.
e) Obtenga el CV(%) del ensayo e interprételo.
f) Realice la prueba de comparación de medias usando el criterio propuesto por Tukey, para la interacción especies × distanciamientos.
g) Lea el artículo original y haga un resumen de la importancia de este estudio y explique el manejo del experimento.

7. En el municipio de Asunción Mita, Jutiapa, durante la época seca de 1980-81 se realizó un experimento con el propósito de evaluar la respuesta de la variedad Chata mexicana de cebolla, cultivada bajo riego a la fertilización con nitrógeno y fósforo, así como establecer niveles óptimos de tales elementos, tanto desde el punto de vista económico como fisiológico.

   En el experimento se incluyeron 12 tratamientos, provenientes de 4 niveles de N (0, 50, 100 y 150 kg/ha) y de 3 de P_2O_5 (0, 25 y 50 kg/ha) en un diseño en bloques al azar con cuatro repeticiones. La distancia de siembra utilizada fue de 0.15 × 0.15 m. La totalidad del fósforo y el 50% del nitrógeno fueron aplicados 10 días después del trasplante y el 50% restante de nitrógeno 25 días después de la primera aplicación.

   Al momento de la cosecha se obtuvo el rendimiento de bulbo de la primera, segunda y tercera calidad (en función del tamaño), cuyo totales se presentan en el Cuadro 1.

Cuadro 1. Rendimiento total (tm/ha) de bulbo de cebolla Chata mexicana obtenido en el experimento sobre fertilización con nitrógeno y fósforo en Asunción Mita, Jutiapa (1980-81).

Tratamiento	N	P_2O_5	Bloque I	II	III	IV	Total	Media
1	0	0	12.764	10.522	11.358	16.03	50.674	12.6685
2	0	25	13.941	5.698	10.674	14.131	44.444	11.111
3	0	50	9.839	9.117	9.459	9.117	37.532	9.383
4	50	0	25.337	16.182	20.323	26.363	88.205	22.05125
5	50	25	19.829	18.613	21.197	27.236	86.875	21.71875
6	50	50	28.262	17.056	20.171	32.213	97.702	24.4255
7	100	0	19.297	25.831	28.756	25.261	99.145	24.78625
8	100	25	24.311	22.222	29.136	29.63	105.299	26.32475
9	100	50	26.553	19.829	21.349	28.604	96.335	24.08375
10	150	0	32.403	24.653	26.363	26.553	109.972	27.493
11	150	25	33.086	19.639	28.946	27.578	109.249	27.31225
12	150	50	25.337	20.855	25.337	30.313	101.842	25.4605
	Total		270.959	210.217	253.069	293.029	1027.274	21.40154

Fuente: Asabá Rivas, R.A. Niveles de nitrógeno y fósforo en el rendimiento y calidad de la cebolla (*Allium cepa* L.) en el Valle de Asunción Mita, Jutiapa. Tesis Ing. Agr. USAC. Facultad de Agronomía.

1. Hipótesis estadísticas

   Ho: $\forall ij = 1, 2, \ldots, 12\ \tau_i = \tau$
   Ha: $i/\tau_i \neq \tau$

   El efecto de los tratamientos puede descomponerse en:

   H_{01}: No existe diferencia significativa en el rendimiento producido por los diferentes niveles de nitrógeno.

H_{02}: No existe diferencia significativa en el rendimiento producido por los diferentes niveles de fósforo.

H_{03}: No existe interacción entre los diferentes niveles de nitrógeno y de fósforo en la fertilización de cebolla.

Las hipótesis alternativas son las que plantean en cada caso la existencia de diferencias significativas en los efectos principales y la interacción correspondiente.

2. Modelo estadístico-matemático

$$Y_{ijk} = \mu + N_i + P_j + (NP)_{ij} + \beta_k + \varepsilon_{ijk}$$

i = 1, 2, 3, 4 = a (niveles de nitrógeno)

j = 1, 2, 3 = b (niveles de fósforo)

k =1, 2, 3, 4 = r (bloques o repeticiones)

Y_{ijk} = rendimiento total (tm/ha) de bulbo obtenido en el k-ésimo bloque, j-ésimo nivel de P_2O_5 y en el i-ésimo nivel de N.

μ = media general del rendimiento

N_i = efecto del i-ésimo nivel de N

P_j = efecto del j-ésimo nivel de P_2O_5

$(NP)_{ij}$ = interacción entre N y P_2O_5

β_k = efecto del k-ésimo bloque o repetición

ε_{ijk} = error experimental asociado a la ijk-ésima unidad experimental.

Cuadro 2 Totales del rendimiento obtenido de acuerdo a la combinación N,P

Nitrógeno	Fósforo			Total	Media
	0	25	50		
0	50.674	44.444	37.532	132.65	11.05417
50	88.205	86.875	97.702	272.782	22.73183
100	99.145	105.299	96.335	300.779	25.06492
150	109.972	109.249	101.842	321.063	26.75525
Total	347.996	345.867	333.411	1027.274	
Media	21.74975	21.61669	20.83819		

Analice la gráfica de las medias del rendimiento por cada interacción, esto es una indicación del grado de interacción y del tipo de respuesta que existe para ajustar un modelo de regresión.

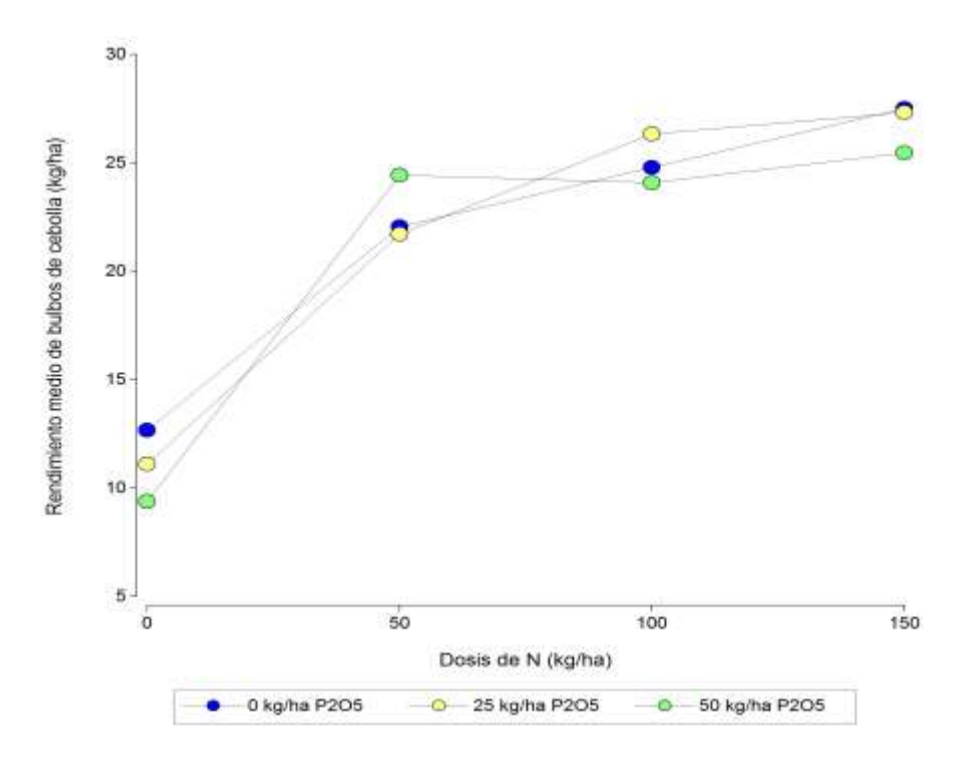


Figura 1 Respuesta del cultivo de la cebolla a la fertilización con nitrógeno y fósforo en Asunción Mita, Jutiapa (1980-1981).

Realice lo siguiente:

1. El análisis de varianza, considerando un nivel de 5% de significancia

2. En experimentos como el anterior, en los cuales los niveles de los factores que se estudian son cuantitativos, generalmente no se recomienda realizar prueba de comparación múltiple de medias (Tukey, Duncan, SNK, etc.), debido a que éstas solo permiten hacer comparaciones entre los niveles evaluados. La técnica que debe emplearse en estos casos, para obtener niveles óptimos (ya sea de carácter fisiológico o económico) es el análisis de regresión entre los niveles de los insumos y las medias del rendimiento obtenido. Por lo tanto:

   2.1 Estime el mejor modelo de regresión para explicar el rendimiento de la cebolla en función de la dosis de nitrógeno.

   2.2 ¿Cuál es el máximo rendimiento posible de obtener? ¿Con qué dosis de nitrógeno? (esta dosis es la llamada el óptimo fisiológico).

   2.3 Estime la función de ganancia y con ella encuentre la máxima ganancia que es posible obtener y la dosis de nitrógeno que se obtiene (esta dosis es la llamada el óptimo económico)

8. Mazariegos, L. (2011) realizó el trabajo de tesis titulado: Efecto de cuatro concentraciones de ácido indolbutirico (IBA) y tres niveles de consistencia de estacas en la propagación asexual de papausa (*Annona diversifolia* Saff; Anonaceae). La investigación fue realizada en el municipio de Pajapita, San Marcos, utilizando un diseño completamente al azar con arreglo combinatorio y cuatro repeticiones, la unidad experimental estuvo formada por 10 estacas de papausas variedad Pajapita, colocadas cada una en bolsas de 22.86 cm de diámetro × 30.48 cm de altura, con sustrato preparado con una mezcla de: 50% de tierra, 30% de materia orgánica y 20% de piedra pómez. A continuación se presentan los resultados de la medición de la altura (en mm) del rebrote de estacas de papausa, 90 días después de la siembra.

			Repeticiones			
Tratamiento	Tipo de estaca	IBA	I	II	III	IV
T1	Herbácea	0	41	73	82	87
T2	Herbácea	1000	71	42	66	73
T3	Herbácea	2000	58	79	85	71
T4	Herbácea	3000	55	74	65	54
T5	Semileñosa	0	114	72	76	86
T6	Semileñosa	1000	102	61	95	90
T7	Semileñosa	2000	39	67	109	65
T8	Semileñosa	3000	86	90	83	93
T9	Leñosa	0	54	80	86	98
T10	Leñosa	1000	149	82	36	45
T11	Leñosa	2000	39	22	34	42
T12	Leñosa	3000	134	70	69	79

Fuente: Mazariegos Pérez, L.A. Efecto de cuatro concentraciones de acido indolbutirico (IBA) y tres niveles de consistencia de estacas en la propagación asexual de papausa (*Annona diversifolia* Saff; Anonaceae). Tesis Ing. Agr. Universidad Rafael Landívar. Facultad de Ciencias Ambientales y Agrícolas. 53 p. Disponible en: http://biblio3.url.edu.gt/Tesario/2011/06/17/Mazariegos-Leonel.pdf

a) Describa el modelo estadístico matemático asociado a este experimento.
b) Construya un gráfico para estudiar la posible interacción entre los niveles de los factores.
c) Realice el análisis de varianza.
d) Verifique los supuestos del modelo estadístico matemático.
e) En caso de ser necesario realice una prueba de comparación múltiple de medias.
f) Concluya con base en los resultados.

8.3 EXPERIMENTOS FACTORIALES CON ARREGLO EN PARCELAS DIVIDIDAS

8.3.1 Introducción

Los experimentos en parcelas divididas (*split-plot design*, en inglés) así como los experimentos en arreglo combinatorio, incluyen todas las posibles combinaciones de dos ó más factores en sus diferentes niveles, y la diferencia entre esos tipos de arreglos está en la manera de instalación de los experimentos y en el esquema del análisis de varianza.

Si se considera un diseño en bloques al azar, con **r** bloques y **a** niveles del factor A. Si cada una de las **ra** parcelas se divide en **b** subparcelas, y entre esas **b** subparcelas se distribuyen al azar los **b** niveles de un factor B, se tiene entonces, una generalización del diseño en bloques al azar, conocida como arreglo en parcelas divididas.

Este tipo de arreglo es útil cuando ciertos niveles de un factor A, para que sean aplicados en el experimento, requieren grandes parcelas, como ocurre en los sistemas de riego, los distanciamientos entre surcos, etc.; y además, esos niveles serán combinados con los niveles del factor B, pudiendo ser: fertilizantes, variedades, etc.

Los niveles del factor A son distribuidos entre las parcelas grandes, las cuales sufrirán una división, de tal modo que los niveles del factor B, que no requieren grandes parcelas, puedan ser distribuidos entre las subparcelas (parcelas pequeñas). De esta manera, se crean dos estructuras, una estructura a nivel de parcelas grandes, con los niveles del factor A, y otra estructura a nivel de subparcelas dentro de cada parcela grande, con los niveles del factor B.

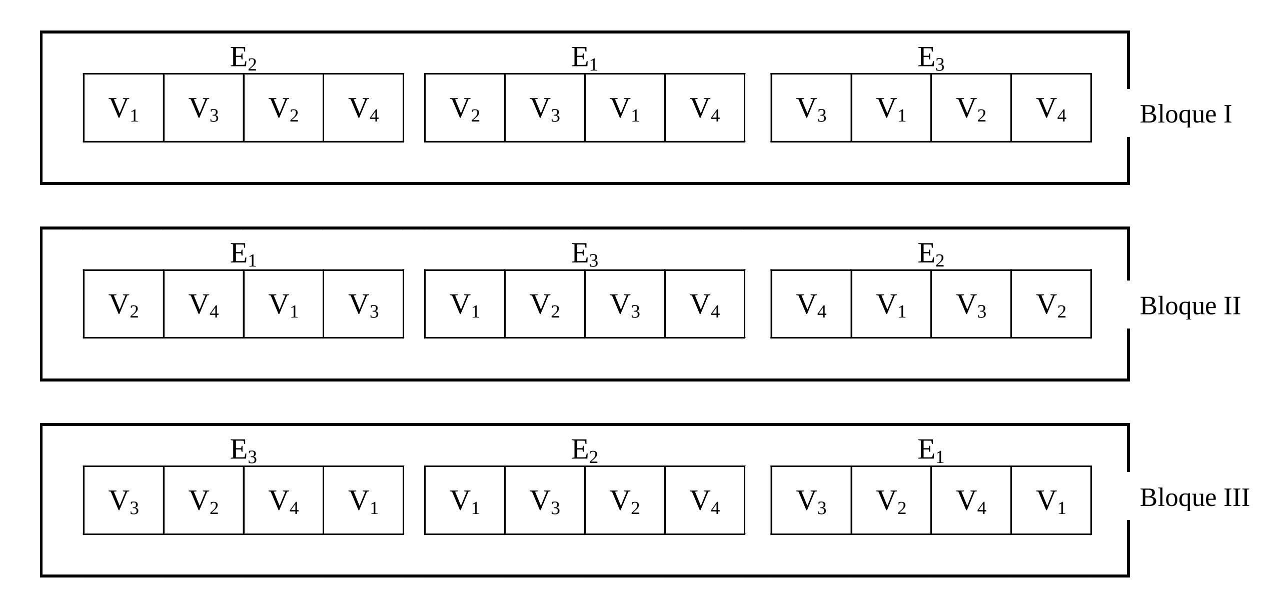
Los niveles aplicados en las parcelas grandes son denominados: **tratamientos primarios**, y los niveles del factor aplicado en las parcelas pequeñas son denominados: **tratamientos secundarios**. Debido a esta estructura, los tratamientos primarios son confundidos con las parcelas grandes, en tanto que los tratamientos secundarios no son confundidos, por eso se debe asignar el factor de mayor interés, en la medida de lo posible, a las parcelas pequeñas. Este tipo de arreglo, es muchas veces preferido (en comparación con el arreglo combinatorio), debido a las facilidades de instalación de los tratamientos en el área experimental.

Desde el punto de vista estadístico, los experimentos que utilizan arreglo combinatorio, por lo general, son más eficientes que los experimentos donde se utiliza arreglo en parcelas divididas, puesto que, en tanto en el arreglo combinatorio existe un único residuo para el cálculo de F y en las comparaciones múltiples, en los experimentos donde se utiliza arreglo en parcelas divididas hay dos residuos, uno referente a parcelas grandes y otro referente a parcelas pequeñas, siendo que el residuo de parcelas grandes, generalmente tiene un número de grados de libertad pequeño, llevando a poca sensibilidad en el análisis.

Se pueden distinguir dos tipos de parcelas divididas, de conformidad con la estructura de las parcelas pequeñas, o sea:

i) Parcelas divididas en el espacio

Cuando cada parcela grande es dividida en subáreas, constituyendo las subparcelas (o parcelas pequeñas). Suponiendo, por ejemplo, que se aplica en las parcelas grandes 3 distanciamientos entre surcos, y en las parcelas pequeñas, 4 variedades de caña de azúcar, distribuidos en 3 bloques, cuyo croquis es el siguiente:



Observando el croquis anterior, se verifica que cada parcela grande, independiente del diseño experimental adoptado, constituye un bloque para las parcelas pequeñas.

ii) Parcelas divididas en el tiempo

Las parcelas grandes no son divididas en áreas, pero periódicamente son tomadas muestras de cada parcela, constituyendo cada muestra como una parcela pequeña. Por ejemplo, aplicando en las parcelas grandes reguladores de crecimiento y considerando a nivel de parcelas pequeñas las épocas o periodos, caracterizadas por las muestras retiradas de las unidades experimentales durante un cierto periodo de tiempo, pudiendo ser, por ejemplo, a cada 15 días durante 6 meses.

La estructura de parcelas divididas en el tiempo, es caracterizada cuando se aplica a las parcelas los niveles de un único factor A y se toman medidas repetidas en ocasiones sucesivas sobre la misma parcela, admitiendo que esas medidas tomadas en ocasiones distintas tienen varianzas homogéneas y son igualmente correlacionadas. En el caso de que esas suposiciones no sean satisfechas, se tiene la estructura de medidas repetidas en el tiempo, cuya metodología de análisis de los datos experimentales pasa a ser el análisis multivariado, que adopta una hipótesis más general sobre la estructura de covarianzas entre medidas repetidas tomadas sobre una misma parcela, que es el análisis de perfil.

El objetivo del análisis de ese tipo de experimento es estudiar el efecto global de los niveles del factor A a lo largo del periodo definido, y la variación de esos niveles a lo largo del tiempo.

En ese caso, es interesante observar que la caracterización de las parcelas pequeñas conduce a una sistematización, puesto que no es posible aleatorizarlas dentro de cada parcela. Las parcelas pueden estar dispuestas en cualquier diseño experimental, pero lo usual es utilizar el diseño completamente al azar o en bloques al azar.

8.3.2 Modelo estadístico

Para un experimento bifactorial dispuesto en un diseño en bloques al azar con arreglo en parcelas divididas, el modelo estadístico es el siguiente:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \rho_k + (\alpha\rho)_{ik} + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,r \\ k = 1,2,\ldots,b \end{cases}$$

siendo:

Y_{ijk} = Variable de respuesta medida en la ijk - ésima unidad experimental
μ = Media general
β_j = Efecto del j - ésimo bloque
αi = Efecto del i - ésimo nivel del factor A.
$(\alpha\beta)_{ij}$ = Efecto de la interacción del i-ésimo nivel del factor A con el j - ésimo bloque, que es utilizado como residuo de parcelas grandes y es representado por $error_{(a)}$
ρ_k = Efecto del k - ésimo nivel del factor B
$(\alpha\rho)_{ik}$ = Efecto debido a la interacción del i-ésimo nivel del factor A con el k - ésimo nivel del factor B.
ε_{ijk} = Error experimental asociado a Y_{ijk} , es utilizado como residuo a nivel de parcela pequeña, y es definido como: $Error_{(b)}$

8.3.3 Hipótesis

Ho: $\mu_{1..} = \ldots = \mu_{a..}$, contra
Ha: por lo menos $\mu_{i..} \neq \mu_{i'..}$ para $i \neq i'$;

Ho: $\mu_{..1} = \ldots = \mu_{..k}$ contra
Ha: por lo menos $\mu_{..k} \neq \mu_{..k'}$ para $k \neq k'$;

Ho: $\mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu = 0$, contra
Ha: $\mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu \neq 0$ para algún i y k.

8.3.4 Análisis de varianza

A continuación se muestra la forma de presentar los datos, previo a su análisis.

Tratamiento primario	Tratamiento secundario	BLOQUES			$Y_{i.k}$
		I	II	III	
A_1	B_1	Y_{111}	Y_{121}	Y_{131}	$Y_{1.1}$
	B_2	Y_{112}	Y_{122}	Y_{132}	$Y_{1.2}$
	B_3	Y_{113}	Y_{123}	Y_{133}	$Y_{1.3}$
Total parcela grande		$Y_{11.}$	$Y_{12.}$	$Y_{13.}$	$Y_{1..}$
A_2	B_1	Y_{211}	Y_{221}	Y_{231}	$Y_{2.1}$
	B_2	Y_{212}	Y_{222}	Y_{232}	$Y_{2.2}$
	B_3	Y_{213}	Y_{223}	Y_{233}	$Y_{2.3}$
Total parcela grande		$Y_{21.}$	$Y_{22.}$	$Y_{23.}$	$Y_{2..}$
Total bloques		$Y_{.1.}$	$Y_{.2.}$	$Y_{.3.}$	$Y_{...}$

Tabla adicional: Esta tabla es útil para calcular la suma de cuadrados para ambos factores.

Factor "B"	Factor "A"		Total
	A_1	A_2	
B_1	$Y_{1.1}$	$Y_{2.1}$	$Y_{...1}$
B_2	$Y_{1.2}$	$Y_{2.2}$	$Y_{...2}$
B_3	$Y_{1.3}$	$Y_{2.3}$	$Y_{..3}$
Total	$Y_{1..}$	$Y_{2..}$	$Y...$

Cuadro de Análisis de Varianza

Fuente de variación	Grados de libertad	Suma de Cuadrados	Cuadrados medios	Valor de F
Bloques	r – 1	$\sum_{j=1}^{r}\frac{Y_{.j.}^2}{ab}-\frac{Y_{...}^2}{arb}$		
A	a – 1	$\sum_{i=1}^{a}\frac{Y_{i..}^2}{rb}-\frac{Y_{...}^2}{arb}$	SC_A/GL_A	$CM_A/CM_{(a)}$
Error (a)	(a – 1) (r – 1)	$SC_{subt} - SC_{bloque} - SC_A$	$SC_{(a)}/GL_{(a)}$	
Subtotal		$\sum_{j=1}^{r}\sum_{i=1}^{a}\frac{Y_{ij.}^2}{b}-\frac{Y_{...}^2}{arb}$		
B	b – 1	$\sum_{k=1}^{b}\frac{Y_{..k}^2}{ar}-\frac{Y_{...}^2}{arb}$	SC_B/GL_B	$CM_B/CM_{(b)}$
AB	(a – 1) (b – 1)	$\sum_{j=1}^{r}\sum_{k=1}^{b}\frac{Y_{i.k}^2}{r}-\frac{Y_{...}^2}{arb}-SC_A-SC_B$	SC_{AB}/GL_{AB}	$CM_{AB}/CM_{(b)}$
Error (b)	a (b – 1) (r –1)	$SC_{tot} - Sc_{subt} - SC_B - SC_{AB}$	$SC_{(b)}/GL_{(b)}$	
Total	abr – 1	$\sum_{j=1}^{r}\sum_{i=1}^{a}\sum_{k=1}^{b}Y_{ijk}^2-\frac{Y_{...}^2}{arb}$		

8.3.5 Análisis post-Andeva: prueba de comparación múltiple de medias, de acuerdo con el criterio de Tukey

a) Prueba de Tukey para el factor ubicado en la parcela grande

$$W = q_{(a,gl(a),\alpha)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CM(a)}{rb}}$$

b) Prueba de Tukey para el factor ubicado en la parcela pequeña

$$W = q_{(b,gl(b),\alpha)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CM(b)}{ra}}$$

c) Prueba de Tukey para la interacción "AB"

$$W = q_{(ab,gl(b),\alpha)} * S\bar{x}, \quad S\bar{x} = \sqrt{\frac{CM(b)}{r}}$$

8.3.6 Ejemplo de Aplicación

En un experimento se probaron dos métodos de siembra y tres variedades de arroz. Por razones de facilidad en el manejo del mismo, se decidió utilizar un arreglo en parcelas divididas**,** en donde las parcelas grandes correspondieron a los métodos de siembra y en las parcelas pequeñas se sembraron las tres variedades de arroz (*Oryza sativa* L.).

Los tratamientos se distribuyeron mediante un diseño en bloques completos al azar con 4 repeticiones. La variable de respuesta fue kg/parcela de 25 m^2. A continuación se detallan los factores evaluados.

Métodos de siembra:	S: Secano	I: Inundación	
Variedades:	BB: Blue Bonnett	BP: Belle Patna	CR: Criolla

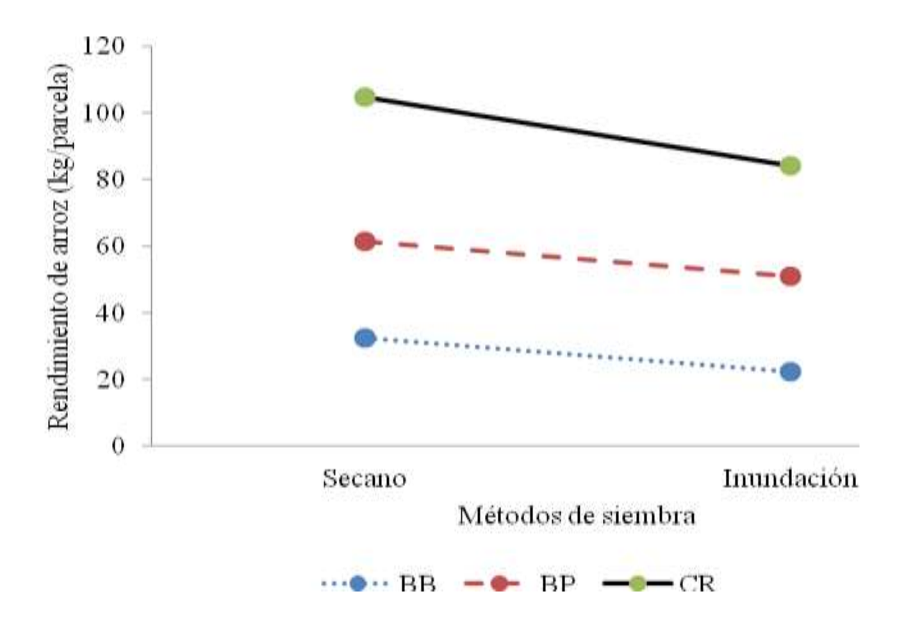
Rendimiento de arroz en kg/parcela de 25 m^2, al comparar distintos métodos de siembra en diferentes variedades de arroz.

Métodos de siembra	Variedades	Bloques			
		I	II	III	IV
S	BB	8.5	8.2	8.0	7.6
	BP	8.0	7.0	7.3	6.9
	CR	10.0	10.8	11.0	11.6
I	BB	5.8	5.6	5.2	5.8
	BP	7.0	7.1	7.3	7.3
	CR	8.3	8.0	8.0	8.7

Tabla adicional:

Variedades	Métodos de siembra		$Y_{..k}$	$\bar{Y}_{..k}$
	Secano	Inundación		
BB	32.30	22.40	54.70	6.84
BP	29.20	28.70	57.90	7.24
CR	43.40	33.00	76.40	9.55
$Y_{i..}$	104.90	84.10		
$\bar{Y}_{i..}$	8.74	7.01		

El gráfico de las interacciones se presenta a continuación:



$$SC_A = \frac{104.9^2 + 84.10^2}{12} - \frac{189^2}{(2)(3)(4)} = 18.03$$

$$SC_{Subtotal} = \frac{26.5^2 + 26^2 + 26.3^2 + 26.1^2 + 21.1^2 + 20.7^2 + 20.5^2 + 21.8^2}{3} - \frac{189^2}{(2)(3)(4)} = 18.41$$

$$SC_{Bloques} = \frac{47.6^2 + 46.7^2 + 46.8^2 + 47.9^2}{(2)(3)} - \frac{189^2}{(2)(3)(4)} = 0.175$$

$$SC_{Error(a)} = 18.41 - 0.175 - 18.03 = 0.205$$

$$SC_B = \frac{54.7^2 + 57.9^2 + 76.4^2}{(2)(4)} - \frac{189^2}{(2)(3)(4)} = 34.308$$

$$SC_{AB} = \frac{32.3^2 + 29.2^2 + 43.4^2 + 22.4^2 + 28.7^2 + 33^2}{4} - \frac{189^2}{(2)(3)(4)} - 18.03 - 34.308 = 7.772$$

$$SC_{Error(b)} = 63.225 - 18.41 - 34.308 - 7.772 = 2.735$$

$$SC_{TOTAL} = 8.5^2 + 8.0^2 + 10^2 + \ldots + 8.7^2 - \frac{189^2}{(2)(3)(4)} = 63.225$$

Cuadro resumen del ANOVA

F.V.	G.L.	S.C.	C.M.	Valor de F	F crítica
Bloques	3	0.175			
A	1	18.03	18.03	265.15*	10.13
Error (a)	3	0.205	0.068		
B	2	34.308	17.154	75.24*	3.89
AB	2	7.772	3.886	17.04*	3.89
Error (b)	12	2.735	0.228		
Total	23	63.225			

$$CV = \frac{\sqrt{0.228}}{7.88} \times 100 = 6.064\%$$

$$F = \frac{CM_A}{CM_{Error(a)}} = \frac{18.03}{0.068} = 265.15$$

$$F = \frac{CM_B}{CM_{Error(b)}} = \frac{17.154}{0.228} = 75.237$$

$$F = \frac{CM_{AB}}{CM_{Error(b)}} = \frac{3.886}{0.228} = 17.04$$

Prueba de comparación múltiple de medias de acuerdo con el criterio de Tukey, para la interacción

a) Medias de las interacciones

Interacción	$\overline{Y}_{ij.}$
A_1B_1	8.08
A_1B_2	7.30
A_1B_3	10.85
A_2B_1	5.60
A_2B_2	7.18
A_2B_3	8.25

b) Comparador

$$Wp = q_{(6,12,0.05)} \times \sqrt{\frac{0.228}{4}} \quad = \quad 4.75 \times 0.239 = 1.134$$

c) Matriz de diferencias

Interacción		A_1B_3	A_2B_3	A_1B_1	A_1B_2	A_2B_2	A_2B_1
	Media	10.85	8.25	8.08	7.30	7.18	5.60
A_2B_1	5.60	5.25*	2.65*	2.48*	1.70*	1.58*	—
A_2B_2	7.18	3.67*	1.07	0.90	0.12	—	
A_1B_2	7.30	3.55*	0.95	0.78	—		
A_1B_1	8.08	2.77*	0.17	—			
A_2B_3	8.25	2.60*	—				
A_1B_3	10.85	—					

d) Presentación de resultados

Interacciones	Producción promedio de arroz	Grupo Tukey
A_1B_3	10.87	a
A_2B_3	8.25	b
A_1B_1	8.08	b
A_1B_2	7.30	b
A_2B_2	7.18	b
A_2B_1	5.60	c

8.3.7 Ejercicios propuestos

1. En un estudio con pasto Napier (*Pennisetum purpureum*) se evaluaron dos métodos de siembra y tres densidades de siembra, bajo un arreglo en parcelas divididas y distribución en bloques al azar. En la parcela grande se ubicó el factor métodos de siembra y en las parcelas pequeñas la densidad (kg.ha^{-1})

   Factor A: Métodos de siembra: M_1= al voleo.
   M_2=20 cm. entre surco

   Factor B: Densidades: D_1=15 kg.ha^{-1}
   D_2=25 kg.ha^{-1}
   D_3=35 kg.ha^{-1}

   La variable de respuesta medida fue el rendimiento expresado en toneladas métricas por hectárea (peso fresco) de pasto. Los resultados obtenidos se presentan a continuación, tal como fueron obtenidos en el campo:

M_1			M_2			
D_1 49.6	D_3 49.9	D_2 50.2	D_2 44.6	D_1 40	D_3 40.2	Bloque I

M_2			M_1			
D_2 45.8	D_1 46.7	D_3 55.3	D_3 42.7	D_1 43.4	D_2 52	Bloque II

M_1			M_2			
D_3 48.7	D_2 44.8	D_1 41.4	D_1 39.2	D_3 39.2	D_2 40	Bloque III

   a) Construya una tabla para resumir la información y realice el ANOVA
   b) Realice la prueba de comparación múltiple de medias (en caso sea necesario)

2. Un experimento fue realizado con el objetivo de evaluar 5 variedades de caña de azúcar (tratamientos principales) y 2 distanciamientos de plantio (tratamientos secundarios), en un diseño bloques al azar con 4 repeticiones. La variable de respuesta medida fue la producción de caña en tm.ha^{-1}.

Variedades	Distanciamientos	Repeticiones			
		I	II	III	IV
V_1	D_1	105.0	101.4	97.2	89.6
	D_2	94.3	91.7	93.5	81.8
V_2	D_1	101.8	96.3	110.0	90.5
	D_2	86.3	90.7	92.4	85.8
V_3	D_1	83.4	60.7	71.3	62.6
	D_2	90.7	58.4	65.2	58.9
V_4	D_1	72.6	54.2	60.5	57.4
	D_2	65.7	56.3	51.3	52.6
V_5	D_1	57.3	50.2	61.3	65.2
	D_2	51.6	53.4	51.9	48.7

a) Describa el modelo estadístico matemático asociado a este experimento
b) Plantee las hipótesis a evaluar.
c) Realice el análisis de varianza
d) Realice, en caso sea necesario, el análisis PostAnova que considere más adecuado.
e) Emita las conclusiones respectivas.

3. Complete la siguiente tabla de ANOVA.

F.V.	G.L.	S.C.	C.M.	Valor de F	F crítica
Bloques	2	52.97			
A	6	28.97			
Error (a)					
B	1	2.72			
AB		9.26			
Error (b)					
Total	41	170.56			

CV = 10.81%, promedio general del experimento = 13.81 toneladas de azúcar/hectárea.

4. López, EA. (1999) evaluó la influencia de la lámina de riego en el efecto del madurante en caña de azúcar *(Saccharum spp.)* variedad CP-722086; en la finca "El Bálsamo", ingenio "Pantaleón". Los datos de producción expresados en toneladas de caña por hectárea (TCH) se presentan en la tabla siguiente; y con los cuales se le solicita realizar el ANOVA correspondiente.

Parcela grande	Parcela pequeña	Repeticiones		
		I	II	III
6/0	Con madurante	119.32	131.44	130.31
	Sin madurante	108.71	116.67	140.91
3/0	Con madurante	134.09	137.12	130.31
	Sin madurante	131.06	144.70	146.22
6/4	Con madurante	123.87	118.94	154.55
	Sin madurante	143.18	119.70	136.37
0/2	Con madurante	115.91	122.35	138.64
	Sin madurante	112.88	114.02	120.46
0/0	Con madurante	131.44	102.27	134.09
	Sin madurante	130.31	122.73	141.29
0/4	Con madurante	117.43	125.76	131.06
	Sin madurante	109.09	113.64	129.93
2/0	Con madurante	118.56	128.79	146.22
	Sin madurante	107.96	142.43	159.09

Referencias:
0/0 = Sin aplicación de riego
2/0 = Distribución de la lámina de riego en 2 riegos previos a la aplicación del madurante y frecuencia de 50 días.
3/0 = Distribución de la lámina de riego en 3 riegos previos a la aplicación del madurante y frecuencia de 25 días.
0/2 = Distribución de la lámina de riego en 2 riegos posteriores a la aplicación del madurante y frecuencia de 25 días.
6/0 = Distribución de la lámina de riego en 6 riegos previos a la aplicación del madurante y frecuencia de 10 días.
6/4 = Distribución de la lámina de riego en 6 riegos previos y 4 posteriores a la aplicación del madurante y frecuencia de 10 días.

5. Se realizó un experimento para conocer la respuesta (cantidad de materia seca MS producida anualmente, $tm.ha^{-1}$) de tres tipos de praderas (Estrella Africana, Guinea y Jaraguá) deterioradas, a las cuales se les aplicaron diferentes prácticas culturales (quema, chapeo y aplicación de herbicida) para su recuperación. El diseño aplicado fue bloques al azar, considerando en la parcela grande el tipo de pradera. Los resultados se presentan a continuación:

Pradera	Prácticas culturales	Repeticiones I	II	III	IV
Estrella Africana	Quema	12.0	13.5	11.0	9.8
	Chapeo	13.5	14.8	10.8	10.0
	Herbicidas	15.5	16.3	12.4	10.5
Guinea	Quema	10.0	10.5	8.7	7.6
	Chapeo	11.0	11.8	9.6	8.3
	Herbicidas	13.0	12.5	11.2	13.1
Jaraguá	Quema	7.5	7.1	6.7	6.3
	Chapeo	8.0	6.8	7.8	7.0
	Herbicidas	9.0	8.3	9.6	7.5

Fuente: Herrera, J.; Barreras, A. 2001.

a) Describa el modelo estadístico matemático asociado a este experimento
b) Plantee las hipótesis a evaluar.
c) Realice el análisis de varianza
d) Realice la prueba de comparación múltiple de medias (en caso sea necesario)
e) Emita las conclusiones respectivas.

6. Batres, De León, Salazar y Saquimux (2000) realizaron una investigación para determinar la mejor edad de los árboles semilleros de pinabete *(Abies guatemalensis)*, así como el período óptimo que debe transcurrir entre la colección y la siembra de la semilla. El estudio se realizó en el paraje Chuisapón de la aldea Chimente del municipio de Totonicapán. Se utilizó un diseño de bloques completos al azar con arreglo en parcelas divididas y cuatro repeticiones. Cada unidad experimental estuvo conformada por 500 semillas sembradas en cada caja germinadora de madera, en donde se utilizó arena de río como sustrato y fue desinfectada con hervida (en ebullición). Los factores evaluados y sus respectivos niveles se citan a continuación:

Factor	**Niveles**
A. Edad de los árboles semilleros o portagranos	A_1. 20 a 40 años
	A_2. 40 a 50 años
	A_3. Mayores de 50 años
B. Fecha de siembra	B_1. Al momento de la extracción de la semilla de los conos.
	B_2. Tres semanas después de la extracción de las semillas.
	B_3. Seis semanas después de la extracción de las semillas.

Los resultados para la variable porcentaje de germinación (datos transformados) se muestran a continuación:

Factor A	Factor B	BLOQUES I	II	III	IV
	B1	2.408	2.324	2.236	2.324
20 a 40 años	B2	2.145	2.324	2.098	2.324
	B3	2.793	2.145	2.569	2.00
	B1	2.757	3.131	2.646	3.098
40 a 50 años	B2	2.408	2.569	2.324	2.098
	B3	2.828	2.79	2.967	2.408
	B1	2.683	3.606	3.606	3.742
> 50 años	B2	2.324	2.683	2.967	3.00
	B3	2.898	2.683	2.236	2.967

En el caso de la variable altura de plántulas de pinabete, expresada en centímetros, los datos obtenidos se muestran a continuación. Para ambas variables realice el ANDEVA y concluya.

Factor A	Factor B	BLOQUES I	II	III	IV
	B1	7.82	6.02	6.70	8.80
20 a 40 años	B2	7.40	7.87	7.50	7.80
	B3	7.50	7.10	7.55	8.35
	B1	7.25	6.90	8.95	8.80
40 a 50 años	B2	8.70	7.85	7.75	6.78
	B3	7.70	7.10	7.90	7.38
	B1	8.10	7.50	9.00	7.60
> 50 años	B2	7.90	7.45	4.20	7.20
	B3	8.00	7.60	8.30	7.90

7. Los datos que se presentan a continuación se refieren a % de brix en el caldo, de un experimento de aplicación de madurantes químicos en caña de azúcar en PLANALSUCAR (Programa Nacional de Melhoramento da Cana-de-Açúcar). El diseño utilizado fue aleatorizado em bloques (8 bloques) con 3 tratamientos. De cada parcela, fueron colectadas muestras de caña después de 0, 2, 4, 6, 8 y 10 semanas de la aplicción de los madurantes. Los resultados se presentan en la tabla siguiente:

		Bloques							
Productos	Semana	I	II	III	IV	V	VI	VII	VIII
Testigo	0	17.7	17.5	17.95	17.9	17.3	17.6	17.4	17.75
	2	16.95	18.6	17.05	18	19.4	18.55	17.8	18.05
	4	18.95	18.1	17.6	18.35	19.2	18.9	18.25	18.7
	6	18.65	18.5	19.65	18.3	20.1	18.75	19.25	17.4
	8	19.7	20.2	19.7	19.7	20.55	20.25	19.8	18.7
	10	20.75	19	19.85	20.8	19.45	21.7	19.1	18.25

Subtotales		**112.7**	**111.9**	**111.8**	**113.05**	**116**	**115.75**	**111.6**	**108.85**
Polaris	0	16.83	16.96	17.25	18.22	17.75	16.9	16.7	17.19
	2	17.7	17.56	18.17	17.72	19.48	18.2	18.58	18.14
	4	18	18.05	18.55	19.52	20.7	18.85	18.41	19.2
	6	18.68	18.83	17.65	20.4	20.03	19.63	18.7	19.71
	8	19.55	19.38	18.1	20.85	20.8	20.42	19.33	21.1
	10	15.56	17.18	19	19.4	20.45	19.92	20	19.65
Subtotales		**106.32**	**107.96**	**108.72**	**116.11**	**119.21**	**113.92**	**111.72**	**114.99**
Ethrel	0	16.63	17.7	17.52	17.52	17.26	17.18	17.43	16.52
	2	17.18	17.75	17.65	17.55	18.61	18.26	18	16.56
	4	18.05	18.2	18.57	19.57	19.03	18.91	18.48	17.73
	6	18.5	19.53	19.08	19.03	19.48	18.95	18.77	16.54
	8	20.65	20.9	18.68	18.55	20.28	20.25	19.27	15.6
	10	19.9	23.96	19.08	19.03	20.62	20.33	20.06	18.85

a) Describa el modelo estadístico matemático asociado a este experimento
b) Plantee las hipótesis a evaluar.
c) Realice el análisis de varianza
d) Realice, en caso sea necesario, el análisis PostAnova que considere más adecuado.
e) Emita las conclusiones respectivas.

8. Un investigador de una compañía de mariscos quiere estudiar el crecimiento bacterial en ostiones y mejillones sujetos a tres temperaturas de almacenamiento. Están disponibles nueve unidades de enfriamiento. Se selecionaron aleatoriamente tres unidades para cada una de las temperaturas. Los ostiones (1) y los mejillones (2) se guardaron por dos semanas en cada una de las unidades de enfriamiento, después de lo cual se contó el número de bacterias en una muestra de ostiones y mejillones. Se registró el logaritmo del conteo bacterial (Y).

Unidad	Temp (ºC)	Marisco	Y
1	0	1	3.6882
1	0	2	0.3505
2	0	1	1.8275
2	0	2	1.7023
3	0	1	5.2327
3	0	2	4.5780
4	5	1	7.1950
4	5	2	5.0169
5	5	1	9.3224
5	5	2	7.9519
6	5	1	7.4195
6	5	2	6.3861
7	10	1	9.7812
7	10	2	10.1352
8	10	1	6.4703
8	10	2	5.0482
9	10	1	9.4442
9	10	2	11.0329

Este es un experimento de parcelas divididas en diseño completamente al azar. El modelo para este experimento es: $y_{ijk} = \mu + T_i + U_{j(i)} + M_k + (TM)_{ik} + \varepsilon_{ijk}$

i = 1, 2, 3 (temperaturas)
j = 1, 2, 3 (repeticiones)
k = 1, 2 (tipo de mariscos)

siendo:

T_i el efecto de la temperatura (aleatorizadas en las parcelas grandes)
$U_{j(i)}$ es el error asociado a la parcela grande (aleatorio)
M_k es el efecto del tipo de marisco (aleatorizado en las parcelas pequeñas)
$(TM)_{ik}$ es el efecto de la interacción entre temperatura × tipo de marisco
ε_{ijk} error asociado a la parcela pequeña (aleatorio)

a) Plantee las hipótesis a evaluar.
b) Realice el análisis de varianza
c) Realice, en caso sea necesario, el análisis PostAnova que considere más adecuado.
d) Emita las conclusiones respectivas.

9. Tavico Laguarca, D.M. (1990) realizó el trabajo de tesis titulado "Evaluación del efecto de cinco momentos de cosecha sobre la calidad molinera de cuatro líneas promisórias de arroz y una variedad de arroz (*Oryza sativa* L.) en el Centro de Producción Agrícola Cristina, Los Amates, Izabal". El diseño utilizado fue bloques completos al azar, con 3 repeticiones, y un esquema en parcelas divididas. En la parcela grande fueron colocados los niveles del factor A, materiales de arroz: A_1= línea IG 2237; A_2= línea CU 4032; A_3= línea IG 2237, A_4= línea IG 2213; A_5= variedad ICTA Polochic. En las parcelas pequeñas fueron aleatorizados los niveles del factor B, momentos de cosecha a partir de la madurez fisiológica del arroz: B_1= 0 días; B_2= 5 días; B_3= 10 días; B_4= 15 días y B_5= 20 días. La unidad experimental estuvo constituida por 6 surcos de 5 m de largo y 0.3 m entre si (5 m × 1.8 m ó 9 m^2). La parcela neta estuvo formada por los 3 surcos centrales.

Dos de las variables medidas fueron:

a) **Rendimiento de grano cosechado:** en los diferentes momentos de corte establecidos, se tomó el peso del arroz en granza (grano con cáscara) obtendio en la parcela neta y se midió el contenido de humedad del grano. Posteriormente se determinó el peso del grano con 12% de humedad (humedad de almacenamiento). El rendimiento de grano se expresó en toneladas métricas por hectárea.

b) **Rendimiento de molino (%):** se determinó el rendimiento de molino, definido como el porcentaje de arroz molinado, entero y quebrado, que se obtiene de la muestra original (1 kg) de arroz en granza. Los resultados se presentan en el cuadro siguiente:

		Rendimento en granza (tm/ha)			Rendimiento en molino (%)		
		Bloques			Bloques		
Líneas	Momentos	I	II	III	I	II	III
A_1	B_1	3.385	4.595	3.720	70	68	70
	B_2	3.317	4.470	3.244	69	66	70
	B_3	2.995	3.526	3.237	68	68	70
	B_4	3.781	3.142	3.104	69	67	68
	B_5	2.938	3.248	2.718	67	66	68
A_2	B_1	3.738	3.202	4.314	64	66	66
	B_2	2.773	3.898	5.122	62	65	59
	B_3	1.954	2.389	4.579	57	61	62
	B_4	1.925	2.599	3.529	54	61	52
	B_5	2.774	1.954	2.646	51	61	52

	B_1	4.396	3.935	4.276	67	66	64
	B_2	4.422	3.797	4.248	66	66	66
A_3	B_3	4.072	3.916	4.100	65	68	67
	B_4	3.356	3.046	3.712	65	64	66
	B_5	3.586	3.600	3.423	65	64	64
	B_1	4.728	4.878	4.519	66	66	64
	B_2	4.173	4.503	4.859	68	66	64
A_4	B_3	4.122	3.947	4.493	65	62	68
	B_4	3.626	3.591	3.469	65	67	65
	B_5	4.283	2.828	2.348	62	66	62
	B_1	3.661	4.014	4.407	67	67	67
	B_2	4.084	4.128	2.696	64	66	64
A_5	B_3	3.908	3.333	3.517	64	65	62
	B_4	2.865	3.093	2.695	62	64	62
	B_5	2.699	2.974	2.901	63	64	62

a) Describa el modelo estadístico matemático
b) Plantee las hipótesis a evaluar
c) Realice el análisis de varianza para cada una de las variables medidas
d) Realice, en caso sea necesario, el análisis PostAnova que considere más adecuado.
e) Emita las conclusiones respectivas.

10. Considere el siguiente conjunto de datos, provenientes de un estudio reportado por Snedecor y Cochran (1971) donde se evaluó el efecto de cuatro fechas del último corte del año (F1, F2, F3 y F4) sobre la productividad de materia seca (ton.acre^{-1}) de tres variedades de alfalfa (*Medicago sativa*): Cossack, Ladak y Ranger. Cada variedad (parcela principal) se ubicó al azar dentro de cada bloque y posteriormente se subdividió aleatoriamente cada variedad en cuatro fechas de corte (subparcela), empleándose seis bloques en total.

Variedad	Fecha	Bloques					
		1	2	3	4	5	6
Ladak	1	2.17	1.88	1.62	2.34	1.58	1.66
	2	1.58	1.26	1.22	1.59	1.25	0.94
	3	2.29	1.60	1.67	1.91	1.39	1.12
	4	2.23	2.01	1.82	2.10	1.66	1.10
Cossack	1	2.33	2.01	1.70	1.78	1.42	1.35
	2	1.38	1.30	1.85	1.09	1.13	1.06
	3	1.86	1.70	1.81	1.54	1.67	0.88
	4	2.27	1.81	2.01	1.40	1.31	1.06
Ranger	1	1.75	1.95	2.13	1.78	1.31	1.30
	2	1.52	1.47	1.80	1.37	1.01	1.31
	3	1.55	1.61	1.82	1.56	1.23	1.13
	4	1.56	1.72	1.99	1.55	1.51	1.33

11. Un estudio fue realizado para evaluar el rendimiento expresado en kilogramos de forraje verde por parcela registrado durante un ciclo de producción de 228 días de ramio (*Boehmeria nivea*), que es una planta textil, forrajera y alimenticia, en función de tres frecuencias de corte (38, 57 y 76 días) y tres niveles de fertilizante (0, 100 y 200 kg/ha). Los niveles de fertilizante se asignaron a las parcelas grandes con base en un diseño de bloques completos al azar (tipos de suelo) con 3 repeticiones; las frecuencias de corte se asignaron a las parcelas pequeñas. Los resultados se presentan a continuación:

Fertilizante	Frecuencia de corte (días)	Bloques		
		I	II	III
0	38	78.90	72.50	78.60
	57	68.10	66.10	69.30
	76	56.90	57.10	53.90
100	38	84.30	99.30	72.90
	57	86.80	108.90	86.60
	76	73.10	73.40	61.70
200	38	95.60	95.20	96.90
	57	97.80	108.10	99.20
	76	90.30	121.40	97.60

Fuente: http://www.unalmed.edu.co/~jarueda/Parcelas.pdf

a) Describa el modelo estadístico-matemático asociado a este diseño y arreglo.
b) Plantee las hipótesis evaluadas
c) Realice el análisis de varianza y concluya
d) En caso de ser necesario aplique la prueba de comparación múltiple de medias.

Consulte:
http://cursosforestalesunamad.blogspot.com/2015/06/diseno-de-parcelas-divididas.html

12. El Proyecto Leña RZ realizó un ensayo sobre especies y fertilización de árboles del género Eucalyptus en la colonia San Rafael, municipio de Guatemala, en un terreno de propiedad municipal. El experimento fue plantado en junio de 1984 a 2.0 × 1.5 m, con planta producida en el vivero Amatitlán del Instituto Nacional Forestal; el terreno estaba cubierto con vegetación natural (pastos y malezas). El diseño utilizado fue de bloques al azar con parcelas divididas. El factor principal lo constituyeron las especies forestales: $\mathbf{A_1}$**.** E. citriodora Hooke vía Brasil, $\mathbf{A_2}$. E. citriodora Hooke Queensland BLSF Lote 1642, $\mathbf{A_3}$. E. robusta, D. E. (Bansefor) y $\mathbf{A_4}$. E. globulus (Bansefor). El factor secundario, ubicado en las parcelas pequeñas, lo constituyeron los niveles de fertilizante: $\mathbf{B_1}$**.** Sin fertilizante, $\mathbf{B_2}$. 10 gr de boro/planta, $\mathbf{B_3}$. 100 gr NPK/planta y $\mathbf{B_4}$. 100 gr NPK/planta + 10 gr de boro/planta. El fertilizante fue aplicado en el fondo del agujero y se cubrió con 5 cm de tierra.

Con base en lo anterior realice lo siguiente:
a) Considere un bloque y muestra cómo sería la aleatorización del experimento.
b) Describa el modelo estadístico matemático asociado a este experimento.
c) Construya una tabla con las fuentes de variación e indique el número de grados de libertad para cada una de ellas.

8.3.8 Programa en SAS para un experimento bifactorial con arreglo en parcelas divididas

8.3.8.1 Diseño: bloques completos al azar

```
OPTIONS nodate nonumber;
DATA parcela;
INPUT bloque metodo $ variedad $ rend;
CARDS;
1   S   BB   8.5
1   S   BP   8
1   S   CR   10
1   I   BB   5.8
1   I   BP   7
1   I   CR   8.3
2   S   BB   8.2
2   S   BP   7
2   S   CR   10.8
2   I   BB   5.6
2   I   BP   7.1
2   I   CR   8
3   S   BB   8
3   S   BP   7.3
3   S   CR   11
3   I   BB   5.2
3   I   BP   7.3
3   I   CR   8
4   S   BB   7.6
4   S   BP   6.9
4   S   CR   11.6
4   I   BB   5.8
4   I   BP   7.3
4   I   CR   8.7
;
PROC GLM;
    CLASS bloque metodo variedad;
    MODEL rend=bloque metodo variedad metodo*variedad bloque*metodo/ss1;
    TEST h=metodo e=bloque*metodo;
    MEANS metodo/TUKEY E=bloque*metodo;
    MEANS variedad/TUKEY
    LSMEANS metodo*variedad/PDIFF E=bloque*metodo;
RUN;
```

8.3.8.2 Diseño: completamente al azar

```
OPTIONS NODATE NONUMBER;
DATA dcaPD;
INPUT agua rep temp resis;
CARDS;
50   1   15   55.78
50   2   15   59.51
50   3   15   61.64
75   1   20   59.56
75   2   20   64.89
75   3   20   58.64
100  1   25   67.53
100  2   25   72.77
100  3   25   66.93
50   1   30   57.00
50   2   30   60.16
50   3   30   63.87
75   1   15   65.26
75   2   15   58.19
75   3   15   59.96
100  1   20   64.71
100  2   20   62.81
100  3   20   62.03
50   1   25   53.69
50   2   25   61.87
50   3   25   54.50
75   1   30   70.96
75   2   30   62.85
75   3   30   64.11
100  1   15   64.40
100  2   15   60.51
100  3   15   62.41
50   1   20   56.30
50   2   20   56.79
50   3   20   51.96
75   1   25   63.37
75   2   25   62.55
75   3   25   61.81
100  1   30   72.59
100  2   30   73.79
100  3   30   74.55
;
PROC GLM;
CLASS agua rep temp;
MODEL resis= agua rep(agua) temp agua*temp;
TEST H=agua E=rep(agua);
run;
```

Ejemplo:

En un ensayo de resistencia de cartón se realizaron preparados de pasta básica con tres distintas cantidades de agua (50, 75 y 100 litros). Cada uno de los preparados (parcelas principales) se repitió tres veces en orden aleatorio a lo largo del tiempo.

Luego, se dividieron los preparados en cuatro fracciones iguales (subparcelas) y se los sometió a distintas temperaturas de cocción (20, 25, 30 y 35 grados), las que fueron asignadas al azar.

La variable en estudio fue la resistencia del cartón obtenido.

Fuente: Balzarini M.G., Gonzalez L., Tablada M., Casanoves F., Di Rienzo J.A., Robledo C.W. (2008). Infostat, Manual del Usuario, Editorial Brujas, Córdoba, Argentina.

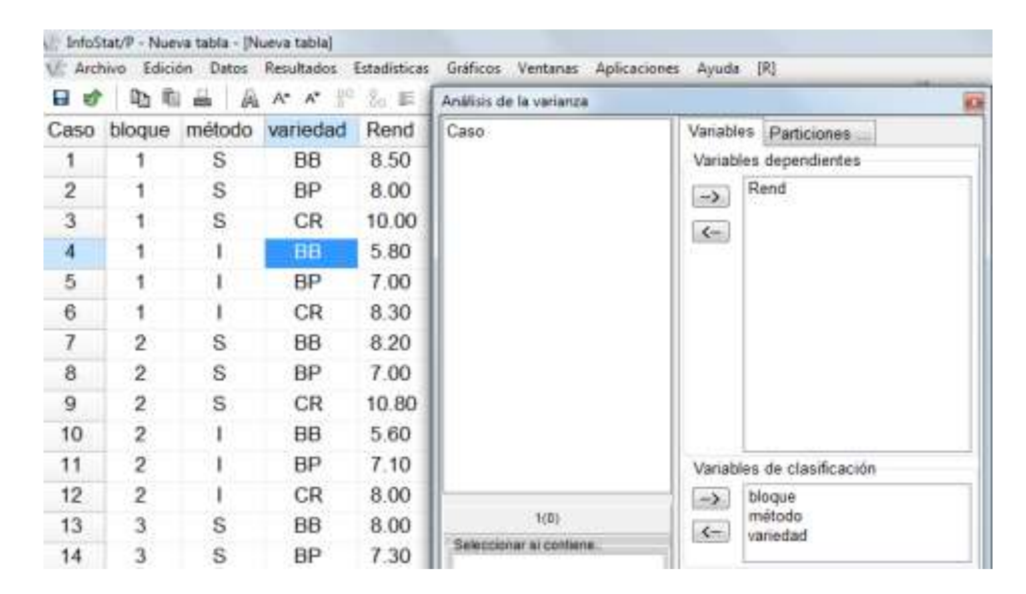


Para resolver el ejercicio de parcelas divididas en un diseño bloques al azar, en Infostat, se presenta en las siguientes ventanas las instrucciones:

1. Luego de ingresar los datos (con cada columna identificada claramente), presione las teclas CONTROL + R, o en el menú ESTADÍSTICAS seleccione la opción ANÁLISIS DE VARIANZA.



2. Especifique las variables, como se muestra a continuación y presione el botón ACEPTAR.



3. En la pestaña MODELO, digite lo siguiente:

   método\método*bloque
   bloque
   método*bloque
   variedad
   método*variedad

4. La salida es la siguiente:

Análisis de la varianza

Variable	N	R^2	R^2 Aj	CV
Rend	24	0.96	0.92	6.06

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor	(Error)
Modelo.	60.49	11	5.50	24.11	<0.0001	
método	18.03	1	18.03	265.97	0.0005	(método*bloque)
bloque	0.18	3	0.06	0.26	0.8558	
método*bloque	0.20	3	0.07	0.30	0.8268	
variedad	34.31	2	17.15	75.22	<0.0001	
método*variedad	7.78	2	3.89	17.05	0.0003	
Error	2.74	12	0.23			
Total	63.23	23				

8.4 EXPERIMENTOS TRIFACTORIALES CON ARREGLO EN PARCELAS SUBDIVIDIDAS

8.4.1 Introducción

En algunas situaciones prácticas, las condiciones experimentales hacen necesario subdividir cada parcela pequeña en diversos niveles de un tercer factor, caracterizándose así, las subpacelas. En ese caso, el ensayo es denominado experimento en parcelas subdivididas o "*split-spli-plot*". Entonces, en este tipo de experimentos, podemos estudiar, simultáneamente, los efectos de tres factores, siendo los tratamientos principales asignados al azar en las parcelas grandes (completamente al azar, en bloques al azar, etc.), los subtratamientos aleatorizados en las parcelas medianas y los subsubtratamientos aleatorizados en las parcelas pequeñas de cada parcela mediana.

Así, por ejemplo, en un experimento en parcelas subdivididas, con los factores distanciamiento (tratamientos principales), frecuencias de riego (tratamientos secundários) e variedades (subsubtratamientos), siendo utilizados: 3 distanciamientos (D_1, D_2 y D_3), 2 frecuencias de riego (F_1 y F_2) y 3 variedades (V_1, V_2 y V_3), si el experimento es instalado en 5 bloques al azar, el croquis de uno de los bloques se presenta a continuación:

D_2		D_1		D_3	
F_1	F_2	F_2	F_1	F_1	F_2
V_2	V_2	V_1	V_2	V_3	V_1
V_1	V_3	V_3	V_3	V_2	V_3
V_3	V_1	V_2	V_1	V_1	V_2

El esquema de análisis de varianza para este experimento es presentado a continuación:

Fuentes de variación	Grados de libertad
Bloques	4
Distanciamientos (D)	2
Residuo (a)	8
(Parcelas)	(14)
Frecuencias de riego (F)	1
Interacción D × F	2
Residuo (b)	12
(Subparcelas)	(29)
Variedades (V)	2
Interacción V × D	4
Interacción V × F	2
Interacción V × D × F	4
Residuo (c)	48
Total	89

Como podrá notar, el experimento en parcelas subdivididas implica tres instancias de aleatorización, por lo tanto para el análisis se deberán tener en cuenta tres errores diferentes: uno para la parcela principal (PP), uno para la subparcela (SP) y otro para las sub-subparcelas. En SAS solo se

deben declarar los errores correspondientes a la PP y a la SP, ya que el tercero de los errores queda declarado por defecto. El error para la parcela principal es la interacción entre bloque y el factor que fue asignado en la PP, en este ejemplo, los distanciamientos. El error para la SP esta dado por la interacción entre el bloque y el factor que esta en la subparcela, en este ejemplo bloque × frecuencias de riego dentro de distanciamientos.

8.4.2 Modelo estadístico-matemático

Para un experimento dispuesto en un diseño en bloques al azar con arreglo en parcelas subdivididas, el modelo estadístico-matemático es el siguiente:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \chi_k + \delta_{ij} + \phi_{ik} + \gamma_{jk} + \eta_{ijk} + b_l + (\alpha b)_{il} + sp_{ijl} + \varepsilon_{ijkl}$$

En la expresión anterior μ representa la media general, α_i el i-ésimo nivel del factor asociado a las parcelas principales, β_j el j-ésimo nivel del factor asociado a las subparcelas dentro de las parcelas principales, χ_k el k-ésimo nivel del factor asociado a las sub-subparcelas (dentro de las subparcelas) y δ_{ij}, ϕ_{ik}, γ_{jk} y η_{ijk} las correspondientes interacciones, b_l el efecto del l-ésimo bloque, ε_{ijkl} el error experimental. Los términos de error son: $(\alpha b)_{il}$ interacción entre el i-ésimo nivel del factor asociado a las parcelas grandes y el l-ésimo bloque (error asociado a las parcelas grandes), sp_{ijl} el error asociado a las parcelas medianas y ε_{ijkl} el error experimental.

8.4.3 Ejemplo

Considere los datos de un experimento em parcelas subdivididas, con 3 fechas de siembra (S_1, S_2 y S_3) del cultivo de remolacha (*Beta vulgaris* L.), como tratamientos principales, 2 niveles de aplicación de insecticidas (I_0 = sin insecticida y I_1= con insecticida), como tratamientos secundários, y 3 épocas de siembra (E_1, E_2 y E_3), como subtratamientos. El experimento fue realizado en 4 bloques, la variable respuesta medida fue la producción por área. Los resultados se presentan a continuación:

Fechas de siembra (S)	Insecticidas (I)	Épocas de Cosecha (E)	Bloques				Totales
			1	2	3	4	
S_1	I_0	E_1	25.7	25.4	23.8	22	96.9
		E_2	31.8	29.5	28.7	26.4	116.4
		E_3	34.6	37.2	29.1	23.7	124.6
	I_1	E_1	27.7	30.3	30.2	33.2	121.4
		E_2	38	40.6	34.6	31	144.2
		E_3	42.1	43.6	44.6	42.7	173.0
S_2	I_0	E_1	28.9	24.7	27.8	23.4	104.8
		E_2	37.5	31.5	31	27.8	127.8
		E_3	38.4	32.5	31.2	29.8	131.9
	I_1	E_1	38	31	29.5	30.7	129.2
		E_2	36.9	31.9	31.5	35.9	136.2
		E_3	44.2	41.6	38.9	37.6	162.3

S_3	I_0	E_1	23.4	24.2	21.2	20.9	89.7
		E_2	25.3	27.7	23.7	24.3	101.0
		E_3	29.8	29.9	24.3	23.8	107.8
	I_1	E_1	20.8	23	25.3	23.1	92.1
		E_2	29	32	26.5	31.2	118.7
		E_3	36.6	37.8	34.8	40.2	149.4
Totales			588.7	574.4	536.6	527.7	2227.4

Fuente: Banzatto y Kronka (2011)

Con estos datos se realizará el análisis de varianza y en caso de ser necesaria una prueba de comparación múltiple de medias, utilizando un nivel de 5% de significancia.

$$\text{F.C.} = \frac{2227.4^2}{(3)(2)(3)(4)} = 68907.09$$

$$SC_{TOTAL} = (25.7^2 + 25.4^2 + \ldots + 40.2^2) - \text{F.C.} = 2840.61$$

$$SC_{BLOQUES} = \frac{1}{(3)(2)(3)}(588.7^2 + 574.4^2 + \ldots + 527.7^2) - \text{F.C.} = 143.46$$

Debemos, inicialmente, organizar el cuadro auxiliar con los totales de cada parcela:

Fechas de siembra (6)	Bloques				Totales S
	1	2	3	4	
S_1	199.9	206.6	191.0	179.0	776.5
S_2	223.9	193.2	189.9	185.2	792.2
S_3	164.9	174.6	163.5	163.5	658.7
Totales de bloques	588.7	574.4	536.6	527.7	2227.4

$$SC_S = \frac{1}{24}(776.5^2 + 792.2^2 + 658.7^2) - \text{F.C.} = 443.69$$

$$SC_{PARCELAS} = \frac{1}{6}(199.9^2 + 206.6^2 + ... + 163.5^2) - \text{F.C.} = 698.90$$

$$SC_{Residuo(a)} = SC_{PARCELAS} - SC_{BLOQUES} - SC_S = 111.75$$

A continuación, organizamos el cuadro auxiliar con los totales de cada subparcela, de la siguiente forma:

Interacciones S × I (3)	Bloques				Totales S × I
	1	2	3	4	
S_1I_0	92.1	92.1	81.6	72.1	337.9
S_1I_1	107.8	114.5	109.4	106.9	438.6
S_2I_0	104.8	88.7	90.0	81.0	364.5
S_2I_1	119.1	104.5	99.9	104.2	427.7
S_3I_0	78.5	81.8	69.2	69.0	298.5
S_3I_1	86.4	92.8	86.5	94.5	360.2
Totales de bloques	588.7	574.4	536.6	527.7	2227.4

$$SC_{SUBPARCELAS} = \frac{1}{3}(92.1^2 + 92.1^2 + ... + 94.5^2) - F.C. = 1524.81$$

En seguida, debemos organizar el cuadro auxiliar para la interacción S × I:

(12)	S_1	S_2	S_3	Totales I
I_0	337.9	364.5	298.5	1000.9
I_1	438.6	427.7	360.2	1226.5
Totales S	776.5	792.2	658.7	2227.4

$$SC_I = \frac{1}{36}(1000.9^2 + 1226.5^2) - F.C. = 706.88$$

$$SC_{S,I} = \frac{1}{12}(337.9^2 + 364.5^2 + ... + 360.2^2) - F.C. = 1191.26$$

$$SC_{S\times I} = SC_{S,I} - SC_S - SC_I = 40.69$$

Entonces: $SC_{Residuo(b)} = SC_{SUBPARCELAS} - SC_{PARCELAS} - SC_I - SC_{S\times I} = 78.34$

A continuación, organizamos el cuadro auxiliar para la interacción E × I:

(12)	E_1	E_2	E_3	Totales I
I_0	291.4	345.2	364.3	1000.9
I_1	342.7	399.1	484.7	1226.5
Totales E	634.1	744.3	849.0	2227.4

$$SC_E = \frac{1}{24}(634.1^2 + 744.3^2 + 849.0^2) - F.C. = 962.34$$

$$SC_{E,I} = \frac{1}{12}(291.4^2 + 345.2^2 + ... + 484.7^2) - F.C. = 1797.05$$

$$SC_{E\times I} = SC_{E,I} - SC_E - SC_I = 127.83$$

Finalmente, debemos organizar el cuadro auxiliar para la interacción E × I:

(8)	E_1	E_2	E_3	Totales S
S_1	218.3	260.6	297.6	776.5
S_2	234.0	264.0	294.2	792.2
S_3	181.8	219.7	257.2	658.7
Totales E	634.1	744.3	849.0	2227.4

$$SC_{S,E} = \frac{1}{8}(218.3^2 + 260.6^2 + \ldots + 257.2^2) - F.C. = 1419.13$$

$$SC_{S\times E} = SC_{S,E} - SC_S - SC_E = 13.10\,.$$

Para la triple interacción, se deben realizar los siguientes cálculos:

$$SC_{S,I,E} = \frac{1}{4}(96.9^2 + 116.4^2 + \ldots + 149.4^2) - F.C. = 2338.55$$

$$SC_{S\times I\times E} = SC_{S,I,E} - SC_S - SC_I - SC_E - SC_{S\times I} - SC_{S\times E} - SC_{I\times E} = 44.02$$

$$SC_{Residuo(c)} = SC_{TOTAL} - SC_{SUBPARCELAS} - SC_E - SC_{S\times E} - SC_{I\times E} - SC_{S\times I\times E} = 168.51$$

El cuadro resumen del ANOVA queda de la siguiente manera:

FV	GL	SC	CM	Valor F
Bloques	3	143.46		
Fechas de Siembra (S)	2	443.69	221.85	11.91**
S×Bloques,residuo (a)	6	111.75	18.63	
Insecticidas (I)	1	706.88	706.88	81.25**
S×I	2	40.69	20.35	2.34
I×Bloques(S),residuo (b)	9	78.34	8.70	
Épocas de cosecha (E)	2	962.34	481.17	102.81**
S×E	4	13.10	3.28	0.70
I×E	2	127.83	63.92	13.66
S×I×E	4	44.02	11.01	2.35
Error, residuo (c)	36	168.51	4.68	
Total	71	2840.61		

Valores de la tabla de F

$F(2,6,0.05) = 5.14$ $F(2,9,0.05) = 4.26$
$F(1,9,0.05) = 5.12$ $F(4,36,0.05) = 2.64$
$F(2,36,0.05) = 3.27$

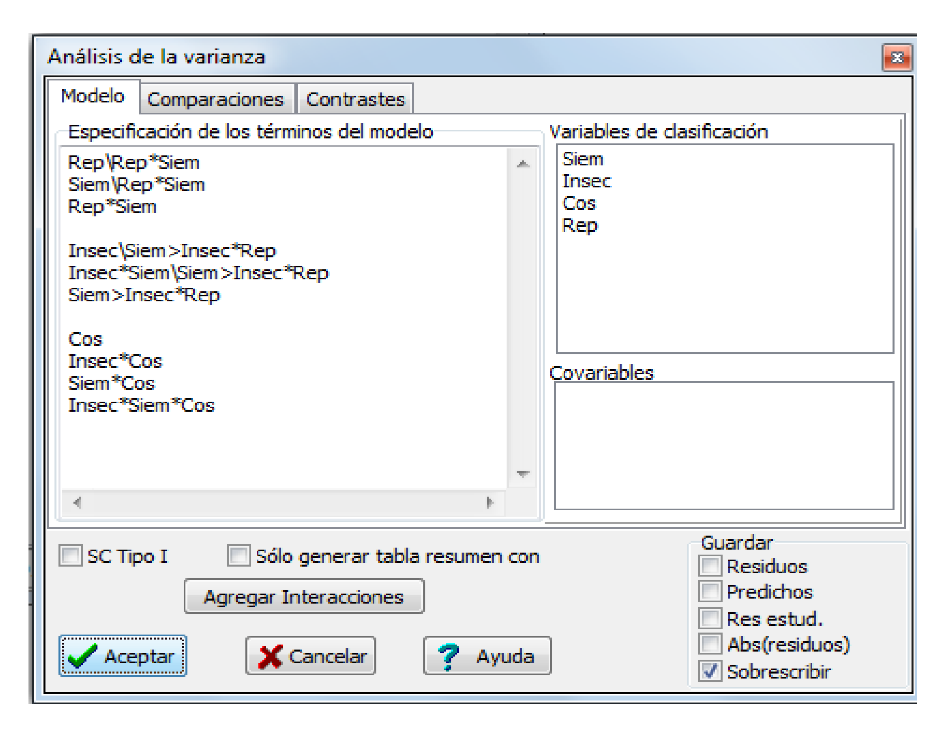
8.4.4 Programa en SAS

```
OPTIONS nodate nonumber;
DATA sub;
INPUT FS $  Ins $ Ecos $ rep $ prod;
DATALINES;
S1     I0     E1     1      25.7
S1     I0     E2     1      31.8
S1     I0     E3     1      34.6
S1     I1     E1     1      27.7
S1     I1     E2     1      38
S1     I1     E3     1      42.1
S2     I0     E1     1      28.9
S2     I0     E2     1      37.5
S2     I0     E3     1      38.4
S2     I1     E1     1      38
S2     I1     E2     1      36.9
S2     I1     E3     1      44.2
S3     I0     E1     1      23.4
S3     I0     E2     1      25.3
S3     I0     E3     1      29.8
S3     I1     E1     1      20.8
S3     I1     E2     1      29
S3     I1     E3     1      36.6
...    ...    ..     ...    ...
...    ...    ...    ...    ...
S1     I0     E1     4      22
S1     I0     E2     4      26.4
S1     I0     E3     4      23.7
S1     I1     E1     4      33.2
S1     I1     E2     4      31
S1     I1     E3     4      42.7
S2     I0     E1     4      23.4
S2     I0     E2     4      27.8
S2     I0     E3     4      29.8
S2     I1     E1     4      30.7
S2     I1     E2     4      35.9
S2     I1     E3     4      37.6
S3     I0     E1     4      20.9
S3     I0     E2     4      24.3
S3     I0     E3     4      23.8
S3     I1     E1     4      23.1
S3     I1     E2     4      31.2
S3     I1     E3     4      40.2
;
PROC GLM;
  CLASS FS Ins Ecos rep;
  MODEL prod = rep FS FS*rep
                   Ins Ins*FS rep*Ins(FS)
                   Ecos Ecos*FS Ecos*Ins Ecos*FS*Ins/ss3;
TEST H = rep FS  E = FS*rep;
TEST H = Ins Ins*FS  E = rep*Ins(FS);
MEANS FS/tukey E=FS*rep;
LSMEANS Ins*Ecos/slice=Ins adjust=tukey PDIFF=all;
LSMEANS Ins*Ecos/slice=Ecos adjust=tukey PDIFF=all;
RUN;
```

Para resolver en Infostat el ejercicio anterior, recuerde que el arreglo en parcelas subdivididas implica tres instancias de aleatorización, por lo tanto para el análisis se deberán tener en cuenta tres errores diferentes: uno para la parcela principal (PP), uno para la subparcela (SP) y otro para las sub-subparcelas (SSP).

En InfoStat solo se deben declarar los errores correspondientes a la PP y a la SP, ya que el tercero de los errores queda declarado por defecto. El error para la parcela principal es la interacción entre bloque y el factor que fue asignado en la PP, en este ejemplo, la época de siembra. El error para la SP esta dado por la interacción entre el bloque y el factor que esta en la subparcela, en este ejemplo Rep*Insec, más la interacción triple de los factores de Bloque, PP y SP, en este caso Rep*Siem*Insec. Esta suma puede reemplazarse en InfoStat por Siem>Insec*Rep (Rep*Siem + Rep*Siem*Insec =Siem>Insec*Rep).

Luego, en la solapa Modelo del menú Análisis de Varianza se deberá escribir lo siguiente:



En esta ventana se ha dejado un espacio entre los términos del modelo para PP, SP y SSP respectivamente para facilitar su visualización. Al oprimir el botón Aceptar se obtendrá el siguiente resultado:

F.V.	SC	gl	CM	F	p-valor	(Error)
Rep	143.23	3	47.74	2.56	0.1505	(Rep*Siem)
Siem	442.99	2	221.50	11.90	0.0082	(Rep*Siem)
Rep*Siem	111.68	6	18.61	3.96	0.0038	
Insec	707.51	1	707.51	81.21	<0.0001	(Siem>Insec*Rep)
Insec*Siem	40.58	2	20.29	2.33	0.1531	(Siem>Insec*Rep)
Siem>Insec*Rep	78.41	9	8.71	1.85	0.0918	
Cos	961.43	2	480.72	102.30	<0.0001	
Insec*Cos	127.63	2	63.82	13.58	<0.0001	
Siem*Cos	13.07	4	3.27	0.70	0.6002	
Insec*Siem*Cos	43.76	4	10.94	2.33	0.0747	
Error	169.17	36	4.70			
Total	2839.47	71				

8.4.5 Ejercicios propuestos

1. López Galindo, H.R. (1984) realizó el trabajo de tesis titulado "Respuesta del arroz (*Oriza sativa* L.) a la aplicación de dos fuentes de 2 fuentes de cal (hidratada y dolomítica), 4 niveles de calcio (0, 500, 1000 y 2000 kg.ha^{-1}) y 3 dosis de fósforo (30, 60 y 90 kg.ha^{-1} de P_2O_5) en suelos ácidos serie Cristina" (Aldea Cristina, Los Amates, Izabal). El diseño experimental utilizado fue el de aleatorizado en bloques (con 4 repeticiones), con esquema en parcelas subdivididas. En las parcelas grandes se aleatorizaron las fuentes de cal, en las medianas los niveles de calcio y en las pequeñas las dosis de fósforo. El tamaño de la parcela bruta fue de 3.0 × 5.0 m (15 m^2), con 10 surcos de 5 m de largo y separados 0.3 m entre sí. El área útil para la toma de datos fue de 1.8 × 5.0 m (9 m^2) correspondiente a los 6 surcos centrales. La siembra se realizó al chorro corrido, utilizando una densidad de 10 grs de semilla por surco (67.5 kg.ha^{-1}) de la variedad de arroz ICTA Virginia. Los rendimientos (kg de grano de arroz con 14% de humedad, por unidad experimental) obtenidos se presentan a continuación:

			Bloques			
Fuente Cal	Nivel Ca	Dosis P_2O_5	I	II	III	IV
Hidratada	0	30	5.5	5.54	5.56	5.37
Hidratada	0	60	6.08	5.86	5.88	5.65
Hidratada	0	90	6.12	6.47	5.78	5.93
Hidratada	500	30	5.16	4.7	4.7	5.03
Hidratada	500	60	5.84	6.72	6.06	5.9
Hidratada	500	90	6.67	6.16	5.24	6.04
Hidratada	1000	30	5.35	5.84	5.93	5.24
Hidratada	1000	60	6.08	5.88	6.59	6.12
Hidratada	1000	90	6.8	6.42	6.12	6.1
Hidratada	2000	30	5.69	5.97	6.27	5.82
Hidratada	2000	60	6.56	6.31	6.33	6.03
Hidratada	2000	90	6.03	6.91	6.25	6.38
Dolomítica	0	30	5.65	4.96	4.92	4.6
Dolomítica	0	60	5.48	5.27	5.14	5.44
Dolomítica	0	90	5.65	5.74	5.71	5
Dolomítica	500	30	5.71	5.74	5.27	5.69
Dolomítica	500	60	5.99	5.71	5.78	5.37
Dolomítica	500	90	6.93	7.29	7.1	7.83
Dolomítica	1000	30	6.06	4.73	6.16	6.03
Dolomítica	1000	60	6.8	5.99	6.16	6.03
Dolomítica	1000	90	6.74	6.38	6.51	7.19
Dolomítica	2000	30	5.48	5.88	5.67	5.69
Dolomítica	2000	60	6.23	6.19	6.51	5.63
Dolomítica	2000	90	6.47	6.29	6.51	6.35

a) Construya el croquis del experimento (realice la aleatorización)
b) Presente el modelo estadístico-matemática y las hipótesis a evaluar.
c) Realice el análisis de varianza y concluya. En caso de ser necesario realice un análisis postANOVA.

2. Sandoval García, C.A. (1987) realizó el trabajo de tesis titulado: "Evaluación de 2 densidades de siembra y 4 niveles de fertilización nitrogenada (urea 46%) en 3 cultivares de bledo (*Amaranthus spp*.) en Villa Canales, Guatemala". Utilizó un diseño de bloques completos al azar, con arreglo en parcelas subdivididas. Los cultivares de bledo se distribuyeron en las parcelas grandes, las densidades de siembra en las parcelas medianas y los niveles de fertilización nitrogenada en las parcelas pequeñas.

Los factores y niveles evaluados fueron:

A. Cultivares de bledo: A_1. 636, A_2. 23-206 y A_3.H.S.
B. Densidades de siembra: B_1. 33,333 y B_2. 83,333 plantas por hectárea
C. Niveles de nitrógeno: C_1.0, C_2.20, C_3.40 y C_4.60 kg.ha^{-1}.

Considere los datos de contenido de proteína en hoja en gramos (% base seca) que se presentan a continuación:

			Bloques		
Cultivares	**Densidad**	**Nitrógeno**	**I**	**II**	**III**
A1	B1	C1	24.24	22.67	22.46
A1	B1	C2	22.96	22.42	22.39
A1	B1	C3	24.56	24.43	24.5
A1	B1	C4	23.24	23.17	23.2
A1	B2	C1	23.79	22.79	23.29
A1	B2	C2	22.64	22.55	22.6
A1	B2	C3	22.94	22.89	22.92
A1	B2	C4	24.66	25.15	24.9
A2	B1	C1	21.15	20.77	20.96
A2	B1	C2	20.7	20.72	20.71
A2	B1	C3	21.12	21.34	21.23
A2	B1	C4	22.37	22.24	22.3
A2	B2	C1	21.88	22.04	21.96
A2	B2	C2	21.91	21.93	21.92
A2	B2	C3	21.39	21.25	21.32
A2	B2	C4	24.14	23.8	23.97
A3	B1	C1	23.64	23.06	23.35
A3	B1	C2	24.44	24.27	24.36
A3	B1	C3	22.77	22.78	22.78
A3	B1	C4	23.76	24.2	23.98
A3	B2	C1	24.86	24.54	24.7
A3	B2	C2	25.01	25.4	25.2
A3	B2	C3	24.08	24.47	24.28
A3	B2	C4	24.17	24.83	24.5

a) Presente el modelo estadístico-matemática y las hipótesis a evaluar.
b) Realice el análisis de varianza y concluya. En caso de ser necesario realice un análisis postANOVA.

8.5 EXPERIMENTOS TRIFACTORIALES EN ARREGLO COMBINATORIO

Para mostrar el análisis e interpretación de resultados de este tipo de experimentos, se utilizarán los datos provenientes del trabajo de investigación de Vargas, VC (2012) que evaluó tres dosis (% de hierro en solución, A_1 = 0.2, A_2 = 0.4 y A_3 =0.6), dos fuentes (B_1= sulfato y B_2= nitrato de hierro) y dos formas de aplicación (C_1 = foliar y C_2 = al tronco) del micronutriente hierro (Fe) en el cultivo de mango (*Mangifera indica* cv Tommy Atkins) en la finca "El Tintero", El Jícaro, departamento de El Progreso.

El diseño experimental utilizado fue de bloques completos al azar, con tres repeticiones y arreglo combinatorio. La unidad experimental estuvo constituida por 2 árboles de mango de doce años de edad, plantados a un distanciamiento de ocho metros entre plantas y ocho metros entre surcos. El experimento fue finalizado al tomar las muestras foliares en el mes de enero, que fue la época de plena floración. La variable de respuesta medida fue la concentración de hierro foliar en la época de floración del cultivo expresado en mg kg^{-1}. El manejo agronómico del cultivo se realizó de acuerdo a las recomendaciones de la finca El Tintero, excepto en cuanto a los tratamientos con micronutrientes que fueron evaluados.

Los resultados obtenidos se presentan en el cuadro siguiente:

% de hierro		Forma de	Repeticiones				
em solución	Fuente	Aplicación	I	II	III	Totales	Medias
0.2	Sulfato	Foliar	125	150	175	450	150.00
0.2	Sulfato	Al tronco	60	100	90	250	83.33
0.2	Nitrato	Foliar	130	145	180	455	151.67
0.2	Nitrato	Al tronco	90	75	80	245	81.67
0.4	Sulfato	Foliar	150	235	250	635	211.67
0.4	Sulfato	Al tronco	105	130	130	365	121.67
0.4	Nitrato	Foliar	150	175	195	520	173.33
0.4	Nitrato	Al tronco	50	75	85	210	70.00
0.6	Sulfato	Foliar	215	415	275	905	301.67
0.6	Sulfato	Al tronco	105	110	100	315	105.00
0.6	Nitrato	Foliar	220	195	205	620	206.67
0.6	Nitrato	Al tronco	175	180	100	455	151.67
		Total bloques	1575	1985	1865	5425	150.69

A partir de este cuadro, se pueden obtener los 3 cuadros siguientes que resumen la información de las interacciones dobles.

	Dosis			
Fuente	0.2	0.4	0.6	Total Fuente
Sulfato	700	1000	1220	2920
Nitrato	700	730	1075	2505
Total Dosis	1400	1730	2295	5425

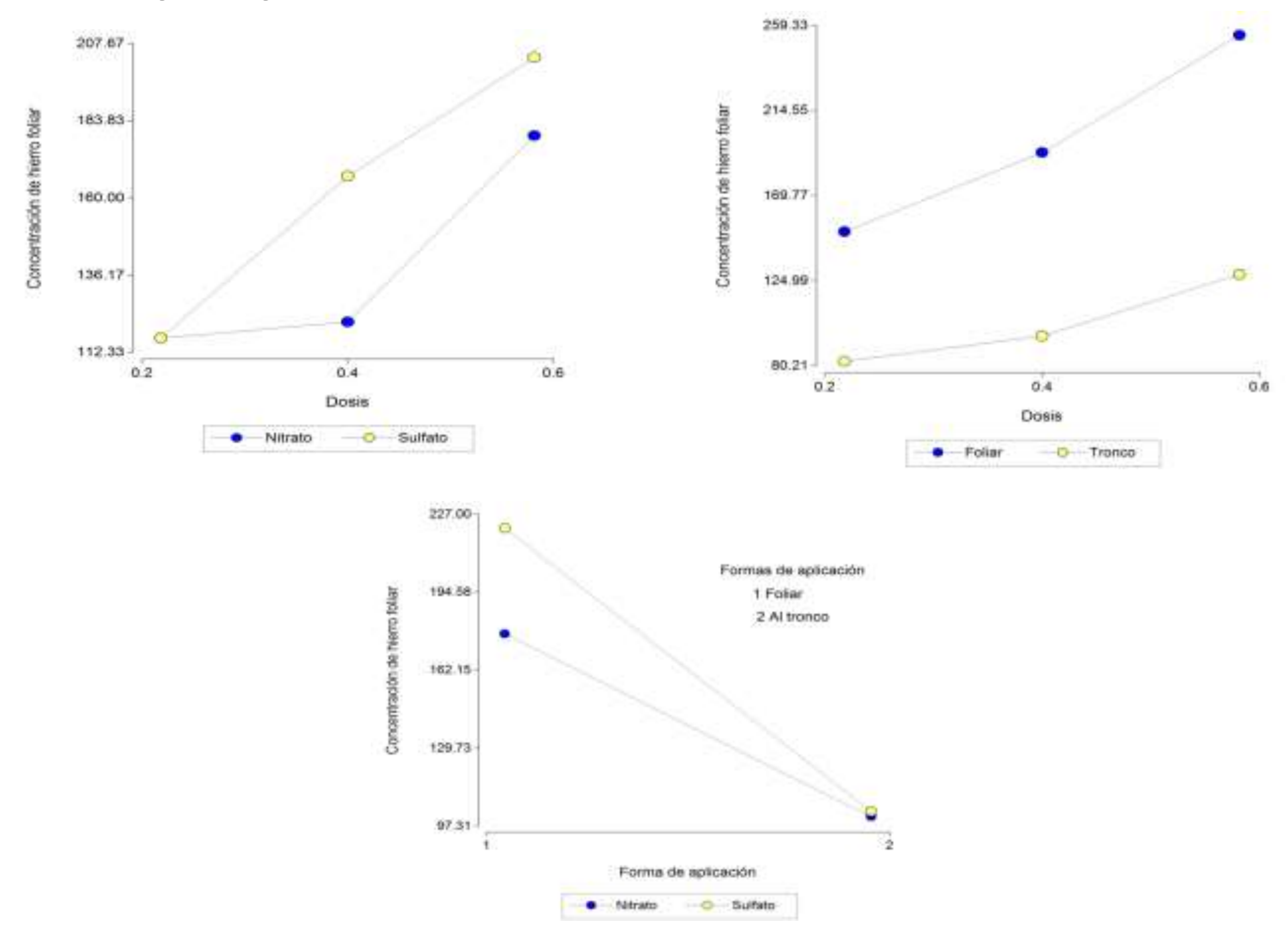
	Dosis			Total
F. Aplicación	0.2	0.4	0.6	F. Aplicación
Foliar	905	1155	1525	3585
Al tronco	495	575	770	1840
Total Dosis	1400	1730	2295	5425

	F. Aplicación		Total
Fuente	Foliar	Al tronco	Fuente
Sulfato	1990	930	2920
Nitrato	1595	910	2505
Total F. Aplición	3585	1840	5425

El modelo estadístico para este diseño es:
$y_{ijkl} = \mu + \tau_i + \delta_j + \gamma_k + (\tau\delta)_{ij} + (\tau\,\gamma)_{ik} + (\delta\gamma)_{jk} + (\tau\delta\gamma)_{ijk} + \beta_l + \varepsilon_{ijkl}$

$i = 1, 2, \cdots, a$ (dosis de hierro, factor A) ;
$j = 1, 2, \cdots, b$ (tipo de fuente de hierro, factor B);
$k = 1, 2, \cdots, c$ (forma de aplicación, factor C);
$l = 1, 2, \cdots, r$ (bloques o repeticiones).

Analice los siguientes gráficos de las interacciones dobles:



Las ecuaciones para realizar el análisis de varianza se presentan a continuación:

F.V	GL	SC	CM
A	a-1	$SC_A = \frac{\sum_{i=1}^{a} y_{i...}^2}{bcr} - \frac{y_{....}^2}{abcr}$	SC_A/GL_A
B	b-1	$SC_B = \frac{\sum_{j=1}^{b} y_{.j..}^2}{acr} - \frac{y_{....}^2}{abcr}$	SC_B/ GL_{AB}
C	c-1	$SC_C = \frac{\sum_{j=1}^{b} y_{..k.}^2}{abr} - \frac{y_{....}^2}{abcr}$	SC_C/ GL_C
AB	(a-1)(b-1)	$SC_{AB} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij..}^2}{cr} - \frac{y_{....}^2}{abcr} - SC_A - SC_B$	SC_{AB}/ GL_{AB}
AC	(a-1)(c-1)	$SC_{AC} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{c} y_{i.k.}^2}{br} - \frac{y_{....}^2}{abcr} - SC_A - SC_C$	SC_{AC}/ GL_{AC}
BC	(b-1)(c-1)	$SC_{BC} = \frac{\sum_{j=1}^{b}\sum_{k=1}^{c} y_{.jk.}^2}{ar} - \frac{y_{....}^2}{abcr} - SC_B - SC_C$	SC_{BC}/ GL_{BC}
ABC	(a-1)(b-1)(c-1)	$SC_{ABC} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r} y_{ijk.}^2}{r} - \frac{y_{....}^2}{abcr} - SC_A - SC_B -$ $SC_C - SC_{AB} - SC_{AC} - SC_{BC}$	SC_{ABC}/ GL_{ABC}
Bloques	(r-1)	$SC_{Bloques} = \frac{\sum_{l=1}^{r} y_{...l}^2}{abc} - \frac{y_{....}^2}{abcr}$	$SC_{BLOQUES}/$ $GL_{BLOQUES}$
Residuo	(abc-1)(r-1)	$SC_{RESIDUO} = SC_{TOTAL} - SC_A - SC_B -$ $SC_C - SC_{AB} - SC_{AC} - SC_{BC} - SC_{ABC} - SC_{BLOQUES}$	$SC_{RESIDUO}/$ $GL_{RESIDUO}$
Total	abcr-1	$SC_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r} y_{ijkl}^2 - \frac{y_{....}^2}{abcr}$	

Bajo este modelo, el objetivo del análisis es realizar los contrastes de hipótesis nula que, junto al estadístico de contraste, se muestran a continuación:

i) $H_{0A} : \tau_1 = ... = \tau_a = 0; F_A = \frac{CM_A}{CM\ Residuo}; Ft_{(a-1),(abc-1)(r-1)}$

ii) $H_{0B} : \delta_1 = ... = \delta_b = 0; F_B = \frac{CM_B}{CM\ Residuo}; Ft_{(b-1),(abc-1)(r-1)}$

iii) $H_{0C}:\gamma_1=...=\gamma_c=0; F_C=\dfrac{CM_C}{CM\,\mathrm{Re}\,siduo}; Ft_{(c-1),(abc-1)(r-1)}$

iv) $H_{0AB}:(\tau\delta)_{ij}=0, \forall_{i,j}; F_{AB}=\dfrac{CM_{(AB)}}{CM\,\mathrm{Re}\,siduo}; Ft_{(a-1)(b-1),(abc-1)(r-1)}$

v) $H_{0AC}:(\tau\gamma)_{ik}=0, \forall_{i,k}; F_{AC}=\dfrac{CM_{(AC)}}{CM\,\mathrm{Re}\,siduo}; Ft_{(a-1)(c-1),(abc-1)(r-1)}$

vi) $H_{0BC}:(\delta\gamma)_{jk}=0, \forall_{j,k}; F_{BC}=\dfrac{CM_{(BC)}}{CM\,\mathrm{Re}\,siduo}; Ft_{(b-1)(c-1),(abc-1)(r-1)}$

vii) $H_{0ABC}:(\tau\delta\gamma)_{ijk}=0, \forall_{i,j,k}; F_{ABC}=\dfrac{CM_{(ABC)}}{CM\,\mathrm{Re}\,siduo}; Ft_{(a-1)(b-1)(c-1),(abc-1)(r-1)}$

Fijado un nivel de significancia α, se rechaza la H_0 correspondiente, si el valor de F > F teórica.

El programa en SAS para analizar este conjunto de datos es el siguiente:

```
*********************************************************************************
PROC glm data = tri;
        CLASS Fue Fe Form Rep;
        MODEL Conc = Fue|Fe|Form Rep;
        lsmeans Fue*Fe*Form/slice=Fue;
        lsmeans Fue*Fe*Form/slice=Fe;
        lsmeans Fue*Fe*Form/slice=Form;
RUN;
*********************************************************************************
```

La salida generada por el programa se presenta a continuación:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	154113.1944	11854.8611	8.84	<.0001
Error	22	29494.4444	1340.6566		
Corrected Total	35	183607.6389			

R-Square	Coeff Var	Root MSE	Conc Mean
0.839362	24.29750	36.61498	150.6944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fue	1	4784.02778	4784.02778	3.57	0.0721
Fe	2	34143.05556	17071.52778	12.73	0.0002
Fue*Fe	2	3043.05556	1521.52778	1.13	0.3396
Form	1	84584.02778	84584.02778	63.09	<.0001
Fue*Form	1	3906.25000	3906.25000	2.91	0.1019
Fe*Form	2	4959.72222	2479.86111	1.85	0.1809
Fue*Fe*Form	2	11287.50000	5643.75000	4.21	0.0283
Rep	2	7405.55556	3702.77778	2.76	0.0851

Diferencias significativas fueron observadas en las dosis de hierro aplicadas (**Fe**), en las formas de aplicación (**Form**) y en la triple interacción (**Fue × Fe × Form**). Siendo esta última la más difícil de interpretar, pues ella puede ser considerada de tres formas: interacción de la interacción **Fue × Fe** con el factor **Form**; interacción de la interacción **Fue × Form** con el factor **Fe**; interacción de la interacción **Fe × Form** con el factor **Fue**. A continuación se presentan cada uno de estos desdoblamientos:

a) Desdoblamiento de la interacción triple para estudiar el comportamiento de las dosis (% de hierro en solución (factor A) en cada combinación de los niveles de Fuente (factor B) y Forma de Aplicación (factor C). Para ello, es necesario construir el cuadro auxiliar siguiente:

A	B_1C_1	B_1C_2	B_2C_1	B_2C_2	Total
A_1	450	250	455	245	1400
A_2	635	365	520	210	1730
A_3	905	315	620	455	2295
Total	1990	930	1595	910	5425

De ese cuadro calculamos:

$$\text{SC A d. } B_1C_1 = \frac{1}{3}(450^2 + 635^2 + 905^2) - \frac{1900^2}{9} = 34905.56$$

$$\text{SC A d. } B_1C_2 = \frac{1}{3}(250^2 + 365^2 + 315^2) - \frac{930^2}{9} = 2216.67$$

$$\text{SC A d. } B_2C_1 = \frac{1}{3}(455^2 + 520^2 + 620^2) - \frac{1595^2}{9} = 4605.56$$

$$\text{SC A d. } B_2C_2 = \frac{1}{3}(245^2 + 210^2 + 455^2) - \frac{910^2}{9} = 11705.56$$

Verificación:

SC A d. B_1C_1 + SC A d. B_1C_2 + SC A d. B_2C_1 + SC A d. B_2C_2 = SC A + SC $_{A\times B}$ + SC $_{A\times C}$ + SC $_{A\times B\times C}$.

El resumen del análisis de varianza para estudiar el comportamiento de las dosis (% de hierro en solución en cada combinación de los niveles de Fuente y Forma de Aplicación se presenta a continuación:

FV	GL	SC	CM	F
A d. B_1C_1	2	34905.56	17452.78	13.02 **
A d. B_1C_2	2	2216.67	1108.33	0.83
A d. B_2C_1	2	4605.56	2302.78	1.72
A d. B_2C_2	2	11705.56	5852.78	4.37 **
Residuo	22	29494.45	1340.66	

Verificamos que existen diferencias entre las dosis de Fe en las combinaciones: Sulfato de Fe y aplicación foliar (B_1C_1) y Nitrato de Fe y aplicación al tronco del árbol (B_2C_2). Para detectar esas diferencias, vamos a aplicar la prueba de Tukey a las medias de dosis en cada combinación de Fuente y forma de aplicación.

%Fe	B_1C_1	B_1C_2	B_2C_1	B_2C_2
A_1	150.00 **b**	83.33 **a**	151.67 **a**	81.67 **a**
A_2	211.67 **b**	121.67 **a**	173.33 **a**	70.00 **b**
A_3	301.67 **a**	105.0 **a**	206.67 **a**	151.67 **a**

a, **b** – en cada columna, medias seguidas de la misma letra no difieren de acuerdo con la prueba de Tukey ($p > 0.05$).

El valor de la diferencia mínima significativa es calculado por la ecuación:

$W = q(t, glee, \alpha) \times \sqrt{\frac{CMee}{r}}$, en que q es obtenido a partir de la tabla de Tukey para los 3 niveles de A y 22 grados de libertad del error experimental (o residuo).

$$W = 3.55 \times \sqrt{\frac{1340.6566}{3}} = 75.04582$$

b) Desdoblamiento de la interacción triple para estudiar el comportamiento de las fuentes de hierro (factor B) en cada combinación de los niveles de dosis (factor A) y Forma de Aplicación (factor C). Para ello, es necesario construir el cuadro auxiliar siguiente:

B	A_1C_1	A_1C_2	A_2C_1	A_2C_2	A_3C_1	A_3C_2	Total
B_1	450	250	635	365	905	315	2920
B_2	455	245	520	210	620	455	2505
Total	905	495	1155	575	1525	770	5425

De ese cuadro calculamos:

$$SC\ B\ d.\ A_1C_1 = \frac{1}{3}(450^2 + 455^2) - \frac{905^2}{6} = 4.17$$

$$SC\ B\ d.\ A_1C_2 = \frac{1}{3}(250^2 + 245^2) - \frac{495^2}{6} = 4.17$$

$$SC\ B\ d.\ A_2C_1 = \frac{1}{3}(635^2 + 520^2) - \frac{1155^2}{6} = 2204.17$$

$$SC\ B\ d.\ A_2C_2 = \frac{1}{3}(365^2 + 210^2) - \frac{575^2}{6} = 4004.17$$

$$\text{SC B d. } A_3C_1 = \frac{1}{3}(905^2 + 620^2) - \frac{1525^2}{6} = 13537.50$$

$$\text{SC B d. } A_3C_2 = \frac{1}{3}(315^2 + 455^2) - \frac{770^2}{6} = 3266.67$$

Verificación:

SC B d. A_1C_1 + SC B d. A_1C_2 + SC B d. A_2C_1 + SC B d. A_2C_2 + SC B d. A_3C_1 + SC B d. A_3C_2 = SC B + SC $_{A\times B}$ + SC $_{B\times C}$ + SC $_{A\times B\times C}$.

El resumen del análisis de varianza para estudiar el comportamiento de las fuentes en cada combinación de los niveles de dosis (% de hierro en solución) y Forma de Aplicación se presenta a continuación:

FV	GL	SC	CM	F
B d. A_1C_1	1	4.17	4.17	0.003
B d. A_1C_2	1	4.17	4.17	0.003
B d. A_2C_1	1	2204.17	2204.17	1.64
B d. A_2C_2	1	4004.17	4004.17	2.99
B d. A_3C_1	1	13537.5	13537.50	10.10 **
B d. A_3C_2	1	3266.67	3266.67	2.44
Residuo	22	29494.45	1340.66	

Verificamos que existe diferencias entre las fuentes de Fe en la combinación: 0.6% de Fe y aplicación foliar (A_3C_1). Para detectar esa diferencia, vamos a aplicar la prueba de Tukey a las medias de fuente en cada combinación de dosis y forma de aplicación.

Fuente	A_1C_1	A_1C_2	A_2C_1	A_2C_2	A_3C_1	A_3C_2
B_1	150.00 **a**	83.33 **a**	211.67 **a**	121.67 **a**	301.67 **a**	105.00 **a**
B_2	151.67 **a**	81.67 **a**	173.33 **a**	70.00 **a**	206.67 **b**	151.67 **a**

a, **b** – en cada columna, medias seguidas de la misma letra no difieren de acuerdo con la prueba de Tukey ($p > 0.05$).

El valor de la diferencia mínima significativa es calculado por la ecuación:

$$W = 2.935 \times \sqrt{\frac{1340.6566}{3}} = 62.045$$

q es obtenido a partir de la tabla de Tukey para los 2 niveles de B y 22 grados de libertad del error experimental (o residuo).

c) Desdoblamiento de la interacción triple para estudiar el comportamiento de las formas de aplicación (factor C) en cada combinación de los niveles de dosis (factor A) y fuente de Fe (factor B). Para ello, es necesario construir el cuadro auxiliar siguiente:

Forma	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_3B_1	A_3B_2	Total
C_1	450	455	635	520	905	620	3585
C_2	250	245	365	210	315	455	1840
Total	700	700	1000	730	1220	1075	5425

De ese cuadro calculamos:

$$\text{SC C d. } A_1B_1 = \frac{1}{3}(450^2 + 250^2) - \frac{700^2}{6} = 6666.67$$

$$\text{SC C d. } A_1B_2 = \frac{1}{3}(455^2 + 245^2) - \frac{700^2}{6} = 7350$$

$$\text{SC C d. } A_2B_1 = \frac{1}{3}(635^2 + 365^2) - \frac{1000^2}{6} = 12150.0$$

$$\text{SC C d. } A_2B_2 = \frac{1}{3}(520^2 + 210^2) - \frac{730^2}{6} = 16016.67$$

$$\text{SC C d. } A_3B_1 = \frac{1}{3}(905^2 + 315^2) - \frac{1220^2}{6} = 58016.67$$

$$\text{SC C d. } A_3B_2 = \frac{1}{3}(620^2 + 455^2) - \frac{1075^2}{6} = 4537.50$$

Verificación:

SC C d. A_1B_1 + SC C d. A_1B_2 + SC C d. A_2B_1 + SC C d. A_2B_2 + SC C d. A_3B_1 + SC C d. A_3B_2 = SC C + SC $_{A\times C}$ + SC $_{B\times C}$ + SC $_{A\times B\times C}$.

El resumen del análisis de varianza para estudiar el comportamiento de las formas de aplicación en cada combinación de los niveles de dosis (% de hierro en solución) y fuentes de Fe se presenta a continuación:

FV	GL	SC	CM	F
C d. A_1B_1	1	6666.67	6666.67	4.97 **
C d. A_1B_2	1	7350.00	7350.00	5.48 **
C d. A_2B_1	1	12150.00	12150.00	9.06 **
C d. A_2B_2	1	16016.67	16016.67	11.95 **
C d. A_3B_1	1	58016.67	58016.67	43.27 **
C d. A_3B_2	1	4537.50	4537.50	3.38
Residuo	22	29494.45	1340.66	

Verificamos que no existe diferencias únicamente entre las formas de aplicación en la combinación: 0.6% de Fe y la fuente nitrato de Fe (A_3B_2). Para detectar las diferencias con las otras combinaciones, vamos a aplicar la prueba de Tukey a las medias de forma de aplicación en cada combinación de dosis y fuente de Fe.

Forma de aplicación	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_3B_1	A_3B_2
C_1	150.00 **a**	151.67 **a**	211.67 **a**	173.33 **a**	301.67 **a**	206.67 **a**
C_2	83.33 **b**	81.67 **b**	121.67 **b**	70.00 **b**	105.00 **b**	151.67 **a**

a, **b** – en cada columna, medias seguidas de la misma letra no difieren de acuerdo con la prueba de Tukey ($p > 0.05$).

El valor de la diferencia mínima significativa es calculado por la ecuación:

$$W = 2.935 \times \sqrt{\frac{1340.6566}{3}} = 62.045$$

q es obtenido a partir de la tabla de Tukey para los 2 niveles de C y 22 grados de libertad del error experimental (o residuo). La forma de aplicación foliar fue la que mostró los mejores resultados.

Observación: De los tres desdoblamientos realizados en la interacción triple, puede ser utilizado cualquiera de ellos, dependiendo apenas del interés del investigador.

Para el caso de un diseño completamente al azar, el modelo estadístico matemático para un experimento trifactorial es el siguiente:

$$y_{ijkl} = \mu + \tau_i + \delta_j + \gamma_k + (\tau\delta)_{ij} + (\tau\gamma)_{ik} + (\delta\gamma)_{jk} + (\tau\delta\gamma)_{ijk} + \varepsilon_{ijkl}$$

$i = 1, 2, \cdots, a$ (niveles del factor A) ;
$j = 1, 2, \cdots, b$ (niveles del factor B);
$k = 1, 2, \cdots, c$ (niveles del factor C);
$l = 1, 2, \cdots, r$ (repeticiones).

Los grados de libertad para el resíduo se calculan así: (abcr−1).

8.5.1 Ejercicios propuestos

1. En Perú fue realizado un experimento en el cultivo de la caña, en caña de segunda soca; se evaluaron 2 dosis de fertilizantes nitrogenados (50 $kg.ha^{-1}$ y 75 $kg.ha^{-1}$), dos variedades (V1- C 8751 y C 1051-73) y dos dosis de Biobrás 16 (40 $ml.ha^{-1}$ y 100 $ml.ha^{-1}$). El Biobrás-16 es un brasinoesteroide de uso amplio en la agricultura teniendo en cuenta su carácter antiestrés y su efecto intensificador en el crecimiento y desarrollo. La variable de respuesta medida fue el rendimiento en toneladas métricas por hectárea de caña, y utilizado el diseño de bloques al azar, con 4 repeticiones. Los resultados se presentan a continuación:

			Bloques			
Dosis N	Variedad	Biobrás	I	II	III	IV
50	V1- C 8751	40	65	68	62	65
50	V1- C 8751	100	58	60	57	56
50	C 1051-73	40	75	72	75	71
50	C 1051-73	100	71	74	72	71
75	V1- C 8751	40	68	65	60	64
75	V1- C 8751	100	70	72	69	68
75	C 1051-73	40	85	86	85	84
75	C 1051-73	100	78	75	79	78

Realice el ANOVA, definiendo claramente el modelo estadístico matemático y las hipótesis a evaluar. En caso de ser necesario aplique una prueba de comparación múltiple de medias.

2. En Perú fue realizado un experimento en el cultivo del plátano con el objetivo de evaluar 3 dosis de mutágeno físico, 2 porcentajes de humedad de la semilla y 2 edades de la semilla. Las dosis del mutágeno físico fueron las siguientes: Rayos Gamma: D_1=100 gy; D_2=200gy y D_3 =300 gy. Los porcientos de humedad fueron: H_1=40 % y H_2=20 %. Las edades de la semilla: E_1=6 meses y E_2=12 meses. Se utilizó un diseño completamente al azar con 4 repeticiones por tratamiento. A continuación se presentan los datos del porcentaje de supervivencia de plantas:

Mutágeno	Humedad	Edad	I	II	III	IV
D_1	H_1	E_1	80.1	82.3	82.5	78.5
D_1	H_1	E_2	75.3	75.6	76.2	75
D_1	H_2	E_1	75.3	74.8	74.5	76.7
D_1	H_2	E_2	73.1	70.3	72.7	71.8
D_2	H_1	E_1	54.6	56.6	54.7	55.1
D_2	H_1	E_2	50.8	51.3	52.6	50.9
D_2	H_2	E_1	74.7	73.8	74.1	72.6
D_2	H_2	E_2	69.5	68.6	70.1	68
D_3	H_1	E_1	64.3	62.6	63	63.8
D_3	H_1	E_2	60.6	61.1	62	61
D_3	H_2	E_1	43.2	45.4	44	44.9
D_3	H_2	E_2	40.8	39.9	40	40.1

Realice el ANOVA, definiendo claramente el modelo estadístico matemático y las hipótesis a evaluar. En caso de ser necesario aplique una prueba de comparación múltiple de medias.

3. Ruiz, E. (1982), realizó su trabajo de tesis en la finca San Rafael, ubicada en el municipio de Mazatenango, Suchitepéquez, instalando un experimento sobre fertilización en piña (*Ananas comosus* L.). Fue utilizado un arreglo combinatorio y un diseño de bloques completos al azar con 3 repeticiones. Los factores y niveles evaluados fueron los siguientes:

A. Nitrógeno: 70, 100 y 130 kg/ha
B. Fósforo: 5, 25 kg/ha
C. Número de aplicaciones: 1, 2 (50% cada una con intervalo de 70 días).

La variable respuesta medida fue el rendimiento de frutos ($tm.ha^{-1}$) obtenido en cada unidad experimental. Los datos se presentan a continuación:

N	P_2O_5	Aplicaciones	Bloques			Total
			I	II	III	
70	5	1	33.303	34.47	35.065	102.838
70	5	2	30.682	33.333	34.091	98.106
70	25	1	32.727	33.838	37.662	104.227
70	25	2	31.818	32.071	33.333	97.222
100	5	1	30.682	32.955	32.517	96.154
100	5	2	32.955	34.375	30.682	98.012
100	25	1	33.523	34.848	30.909	99.28
100	25	2	33.442	33.864	36.818	104.124
130	5	1	31.818	39.015	32.955	103.788
130	5	2	34.091	33.117	30.357	97.565
130	25	1	35.985	35.606	38.131	109.722
130	25	2	32.102	32.386	34.091	98.579
0	0	0	28.409	30.398	27.652	86.459
	Total		421.537	440.276	434.263	1296.076

Realice el ANOVA, definiendo claramente el modelo estadístico matemático y las hipótesis a evaluar. En caso de ser necesario aplique una prueba de comparación múltiple de medias.

Fuente: Ruíz Recinos, E.G. 1982. Evaluación de 3 niveles de N, 2 de P, en dos épocas de aplicación en el cultivo de la piña (*Ananas comosus* Merr.) y sus efectos en el rendimiento bajo condiciones de Mazatenango, Suchitepéquez. Tesis Ing. Agr. Universidad de San Carlos de Guatemala, Facultad de Agronomía. 60 p.

8.6 EXPERIMENTOS FACTORIALES CON TRATAMIENTOS ADICIONALES (TESTIGOS) EN INFOSTAT V. 2016 Y MS EXCEL 2013.

En algunas situaciones prácticas, cuando se trabajan con experimentos factoriales, se presenta la necesidad de incluir uno o dos tratamientos testigos, denominados tratamientos adicionales. Estos testigos son conocidos como: **a)** testigo absoluto (sin aplicación de tratamiento) y **b)** testigo relativo (tratamiento que comúnmente usa el productor). La presencia de estos testigos implica el rompimiento de la estructura factorial del experimento. Por lo que el análisis de varianza sufre algunas modificaciones cuando se incluyen estos tratamientos testigo. A continuación se ilustrará con un ejemplo la resolución de este caso especial, para ello se utilizará Infostat v. 2015 y el complemento de MS Excel: DSAASTAT v. 1.1 (18/03/2011), desarrollado por el profesor Andrea Onofri del Departamento de Agricultura y Ciencias Ambientales de la Universidad de Perugia, Italia y que se encuentra disponible en: http://accounts.unipg.it/~onofri/DSAASTAT/DSAASTAT.htm.

Ejemplo

Considere los datos de incremento de peso (gramos por animal) de cuyes (*Cavia porcellus*) engordados en la fase inicial con concentrados proteicos y energéticos. Los cuyes o cuyos son pequeños roedores herbívoros monogástricos, que se caracterizan por su gran rusticidad, ciclo biológico corto (a los 3 meses alcanzan 1 kg de peso) y buena fertilidad (98%). Estas ventajas han favorecido su explotación y han generalizado su consumo, especialmente en Colombia, Ecuador, Perú y Bolivia.

Los datos a utilizar corresponden a un experimento bifactorial en un diseño completamente al azar con 6 repeticiones. Los factores evaluados (tipos de concentrado) y sus respectivos niveles fueron:

Factor A (concentrados proteicos):

A_1. Harinolina, un subproducto que queda después de que se extrae el aceite a la semilla de algodón. Es decir, es una pasta proteíca de la semilla de algodón.

A_2. Catarina.

Factor B (concentrados energéticos): B_1. Sorgo. B_2. Maíz. B_3. Avena.
Adicionalmente se utilizó un tratamiento testigo, alimentación a base de pasto. Esto hace que se rompa la estructura factorial.

Los resultados del experimento se presentan en el cuadro siguiente (consulte la presentación original disponible en: https://www.youtube.com/watch?v=XjOUVHlIFd8)

Tratamientos	**Proteicos**	**Energéticos**	**Repeticiones**						
			I	**II**	**III**	**IV**	**V**	**VI**	**Promedio**
T1	Harinolina	Sorgo	24.8	25.6	27.8	24.8	24.8	26.7	25.94
T2	Harinolina	Maíz	31.4	31.6	32.8	34.6	34.6	33.9	33.15
T3	Harinolina	Avena	19.8	17.6	20.1	18.6	18.6	18.9	18.93
T4	Catarina	Sorgo	28.3	29.4	27.6	34.6	34.6	32.1	31.10
T5	Catarina	Maíz	35.6	37.1	36.4	34.7	34.7	35.6	35.68
T6	Catarina	Avena	21.4	22.7	23.9	22.9	21.9	22.9	22.45
Pasto			19.4	17.6	18.6	17.6	18.6	19.4	18.53

Caso 1. Diseño completamente al azar en arreglo combinatorio 2 × 3 + 1.

Fuentes de variación	Grados de libertad
Total	41
Proteicos (A)	1
Energéticos (B)	2
A × B	2
A × B vs tratamiento adicional	1
Error experimental	35

Caso 2. Diseño de bloques completos al azar en arreglo combinatorio 2 × 3 + 1.

Fuentes de variación	Grados de libertad
Total	41
Bloques	5
Proteicos (A)	1
Energéticos (B)	2
A × B	2
A × B vs tratamiento adicional	1
Error	30

Observación: en el caso de tener dos testigos, deberá de realizar dos contrastes: **a)** Testigo 1 y testigo 2 vs resto de tratamientos y **b)** Testigo 1 vs testigo 2.

Resumen del procedimiento:

1. Para resolverlo en Infostat v. 2015, no se puede realizar todo el procedimiento, primero hay que analizar los datos como si fuera un experimento simple, esto es considerando los 7 tratamientos (6 interacciones más el tratamiento testigo). De este análisis se tomará la información referente a las siguientes fuentes de variación: Total, A × B vs tratamiento adicional y Error experimental. Caso fuera un diseño de bloques completos al azar, se tomará la información referente a: Total, Bloques, A × B vs tratamiento adicional y Error experimental.

   Organice la tabla de datos de la siguiente manera en Excel y péguela en Infostat v. 2015.

Trats	**Proteicos**	**Energéticos**	**Inc_peso**
T1	Harinolina	Sorgo	24.8
T2	Harinolina	Maíz	31.4
T3	Harinolina	Avena	19.8
T4	Catarina	Sorgo	28.3
T5	Catarina	Maíz	35.6
T6	Catarina	Avena	21.4
T1	Harinolina	Sorgo	25.6
T2	Harinolina	Maíz	31.6
T3	Harinolina	Avena	17.6

T4	Catarina	Sorgo	29.4
T5	Catarina	Maíz	37.1
T6	Catarina	Avena	22.7
T1	Harinolina	Sorgo	27.8
T2	Harinolina	Maíz	32.8
T3	Harinolina	Avena	20.1
T4	Catarina	Sorgo	27.6
T5	Catarina	Maíz	36.4
T6	Catarina	Avena	23.9
T1	Harinolina	Sorgo	24.8
T2	Harinolina	Maíz	34.6
T3	Harinolina	Avena	18.6
T4	Catarina	Sorgo	34.6
T5	Catarina	Maíz	34.7
T6	Catarina	Avena	21.9
T1	Harinolina	Sorgo	24.8
T2	Harinolina	Maíz	34.6
T3	Harinolina	Avena	18.6
T4	Catarina	Sorgo	34.6
T5	Catarina	Maíz	34.7
T6	Catarina	Avena	21.9
T1	Harinolina	Sorgo	26.7
T2	Harinolina	Maíz	33.9
T3	Harinolina	Avena	18.9
T4	Catarina	Sorgo	32.1
T5	Catarina	Maíz	35.6
T6	Catarina	Avena	22.9
Pasto			19.4
Pasto			17.6
Pasto			18.6
Pasto			17.6
Pasto			18.6
Pasto			19.4

Los resultados son los siguientes (la parte en amarillo se tomará para el análisis final)

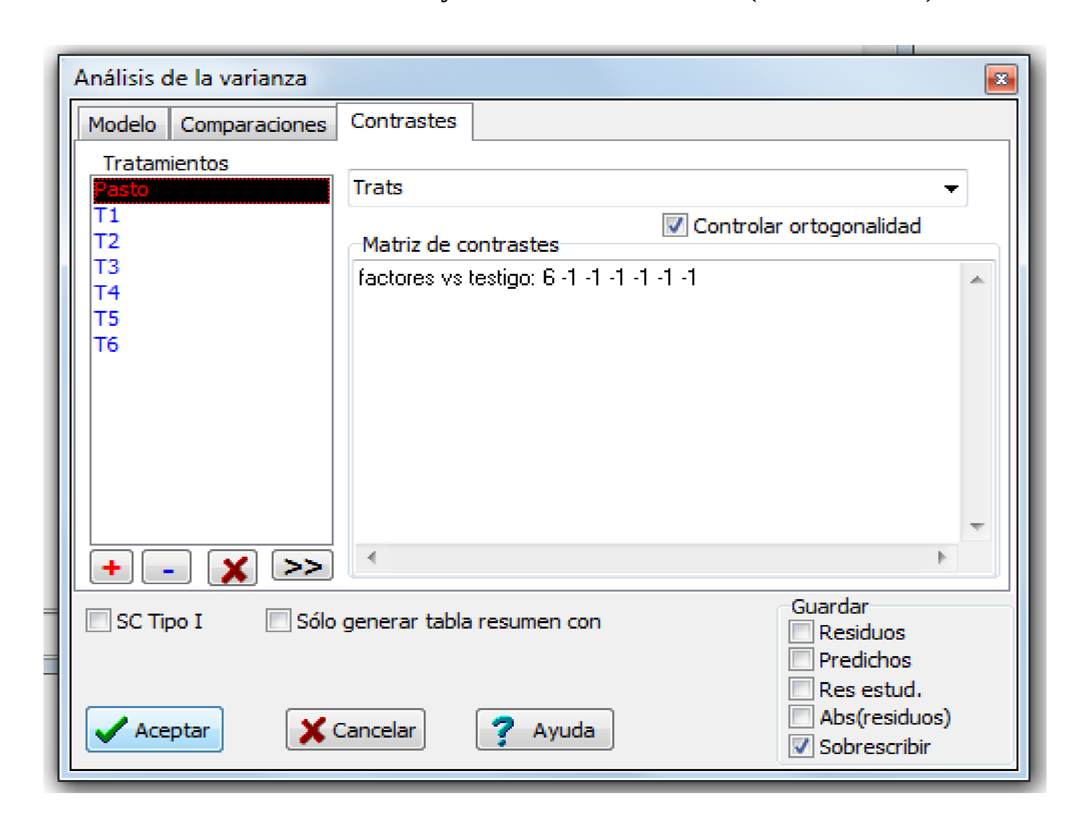
Análisis de la varianza

Variable	N	R²	R² Aj	CV
Inc peso	42	0.95	0.95	5.79

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Trats	1724.41	6	287.40	121.81	<0.0001
Error	82.58	35	2.36		
Total	1806.99	41			

2. Realice el contraste entre Pasto y los tratamientos 1 a 6 (interacciones):

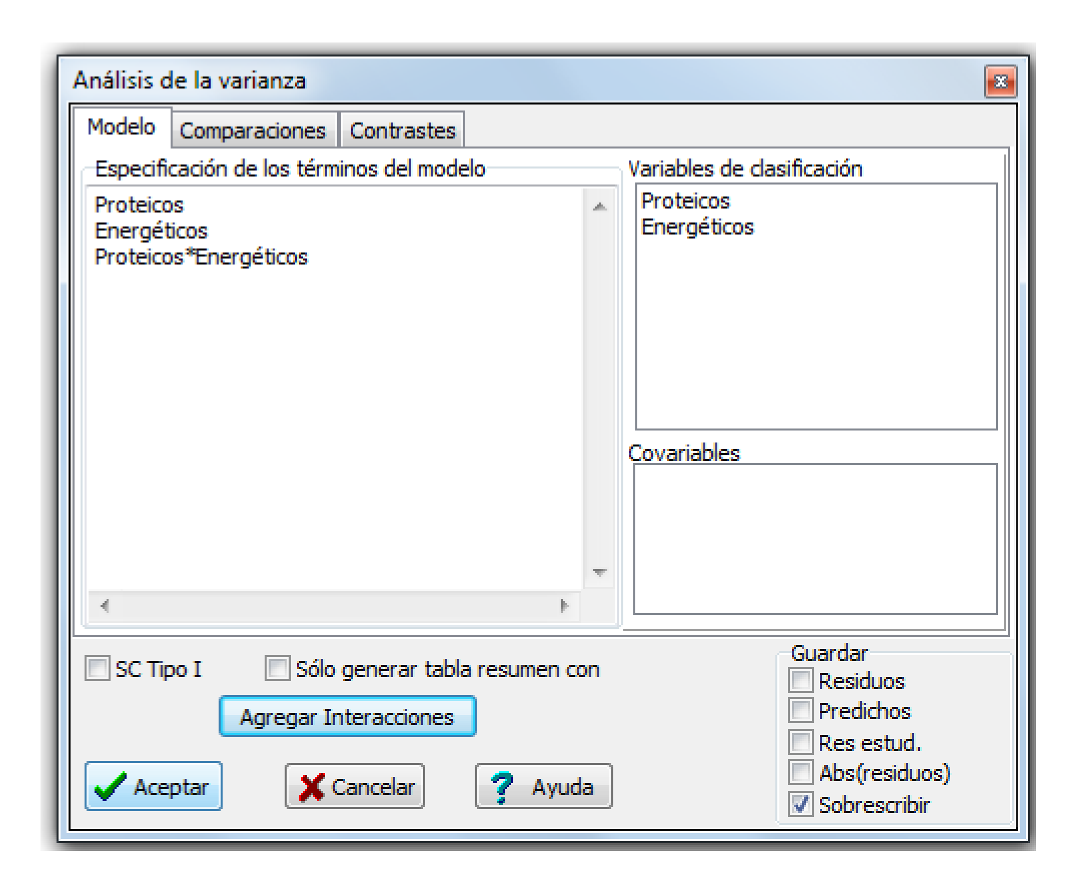


Los resultados son los siguientes (la parte en amarillo se tomará para el análisis final):

Contrastes

Trats	Contraste	E.E.	SC	gl	CM	F	p-valor
Contraste1	-55.87	4.06	445.87	1	445.87	188.97	<0.0001
Total			445.87	1	445.87	188.97	<0.0001

3. Analice los datos en Infostat, respetando la estructura factorial, esto es, sin incluir los datos del (o los) testigo (s). Y tomar la información referente a las siguientes fuentes de variación: Proteicos (A), Energéticos (B) y la interacción A × B. Para ello, de un clic con el botón derecho del ratón, en las filas correspondiente a Pasto y luego seleccione la opción: Desactivar casos seleccionados. Posteriormente proceda a realizar el ANOVA respetando la estructura factorial.



Los resultados son los siguientes (la parte en amarillo se tomará para el análisis final):

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Proteicos	129.96	1	129.96	49.15	<0.0001
Energéticos	1136.32	2	568.16	214.87	<0.0001
Proteicos*Energéticos	12.26	2	6.13	2.32	0.1158
Error	79.33	30	2.64		
Total	1357.87	35			

4. Para unir los dos análisis, se deben de recalcular los valores de F para los efectos principales: Proteicos (A) y Energéticos (B) y para la interacción A × B, así mismo los valores de p. En el caso de los valores de F, se debe de dividir cada Cuadrado Medio (de A, de B y de A × B) entre el Cuadrado Medio del Error. Para calcular el valor de p, use la función de Excel: **=DISTR.F.CD (x, grados_de_libertad1, grados_de_libertad2)**, siendo x el valor de F, grados_de_libertad1 el número de grados de libertad de la fuente de variación y grados_de_libertad2 el número de grados de libertad del error.

Finalmente, se debe dejar la estructura como la mostrada en el Cuadro 1.

F.V.	SC	Gl	CM	F	p-valor
Total	1806.99	41			
Proteicos (A)	129.96	1	129.96	55.07	0.000
Energéticos (B)	1136.32	2	568.16	240.75	0.000
A × B	12.26	2	6.13	2.60	0.116
A × B vs tratamiento adicional	445.87	1	445.87	188.97	<0.0001
Error experimental	82.58	35	2.36		

5. Conclusiones:

a) Presentan diferencias significativas los niveles de los factores principales, concentrados proteicos y energéticos, no así la interacción.

b) Se presenta diferencia entre las combinaciones de concentrados proteicos y energéticos, en comparación con el testigo (alimentación con pasto).

6. Pruebas de comparación múltiple de medias.

Para ello active en Excel el complemento: DSAASTAT y calcule las medias en Infostat, para los factores principales y para los tratamientos T1 a T6 y para el testigo utilice Excel. Los resultados son los siguientes:

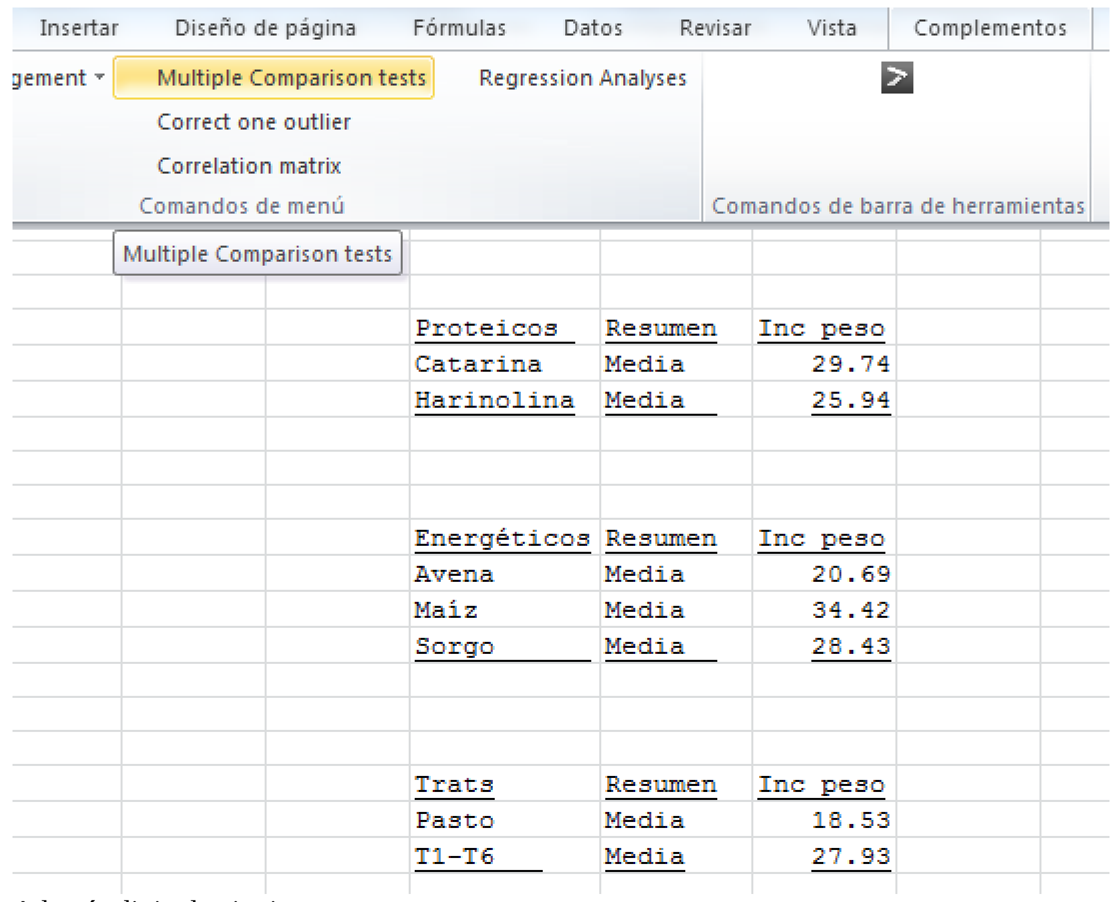
Medidas resumen

Proteicos	Resumen	Inc_peso
Catarina	Media	29.74
Harinolina	Media	25.94

Energéticos	Resumen	Inc_peso
Avena	Media	20.69
Maíz	Media	34.42
Sorgo	Media	28.43

Trats	Resumen	Inc_peso
Pasto	Media	18.53
T1-T6	Media	27.93

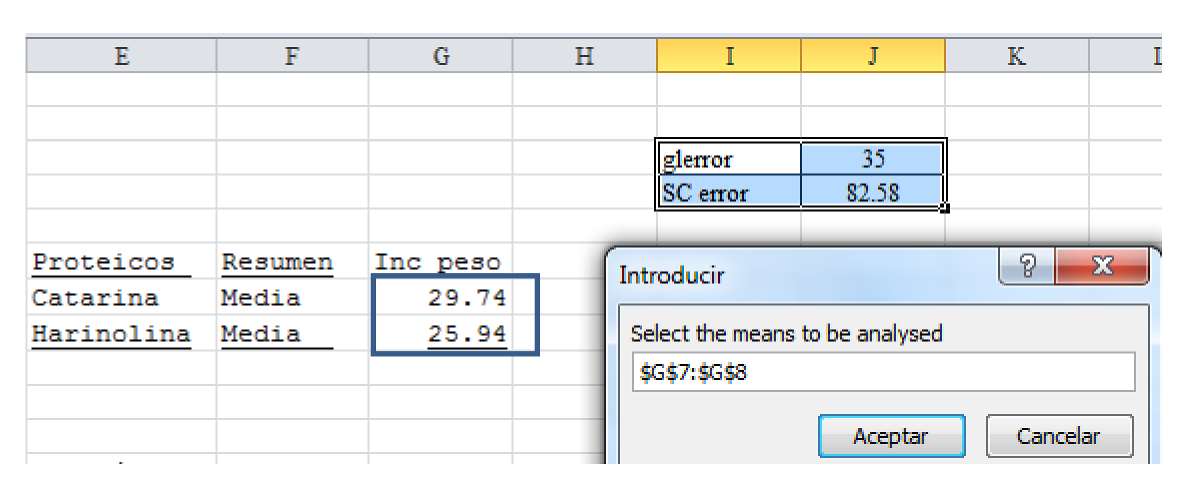
Al instalar el DSAASTAT, aparecerá el siguiente cuadro de diálogo en el menú Complementos de Excel:



Además digite lo siguiente:

glerror	35
SC error	82.58

6.1 Prueba de comparación múltiple de medias para los niveles del factor concentrados proteicos. Luego de dar un clic en Multiple Comparison tests, introduzca las medias y de un clic en Aceptar.



Luego seleccione la prueba de Diferencia Mínima Significativa (DMS) propuesta por Fisher.

Recuerde verificar y en caso necesario alterar la suma de cuadrados residual (o del error experimental, Residual sum of squares), los grados de libertad (Degrees of freedom) y el número de repeticiones (Num. of replications).

MULTIPLE COMPARISON TEST
Procedure: LSD (p= 0.05)
S.E.D.: 0.886834928538671
LSD = 1.80037061947568

1	29.74	a
2	25.94	b

El mejor incremento de peso de los cuyes se obtuvo al utilizar el concentrado proteico Catarina.

6.2 Prueba de comparación múltiple de medias para los niveles del factor concentrados energéticos.

En este caso utilice la prueba de comparación múltiple de medias de acuerdo con el criterio propuestos por Tukey, Tukey HSD method (p= 0.05).

MULTIPLE COMPARISON TEST
Procedure: Tukey HSD method (p= 0.05)
S.E.M.: 0.627086991762782; DF: 35
HSD: 2.17222933946628

2	34.42	a
3	28.43	b
1	20.69	c

El mejor incremento de peso de los cuyes se obtuvo al utilizar el concentrado energético fue maíz.

6.3 Prueba de comparación múltiple de medias LSD de Fisher para la comparación de las medias del testigo contra el promedio de los tratamientos 1 al 6.

MULTIPLE COMPARISON TEST
Procedure: LSD (p= 0.05)
S.E.D.: 0.886834928538671
LSD = 1.80037061947568

2	27.93	a
1	18.53	b

En promedio, el mejor incremento de peso se obtuvo al utilizar los concentrados proteicos y energéticos en comparación con el testigo, alimentación con pasto.

8.6.1 Ejercicios propuestos

1. En una investigación en maíz, utilizando tres niveles de nitrógeno (0, 40 y 80 $kg.ha^{-1}$) y dos de fósforo (0 y 60 $kg.ha^{-1}$) añadiendo un tratamiento adicional (materia orgánica), se utilizó un diseño de bloques completos al azar con tres repeticiones. La variable de respuesta fue kilogramos de grano por parcela neta. Realice el ANOVA y en caso necesario la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey.

			Bloques		
Tratamientos	Nitrógeno	Fósforo	I	II	III
T_1	N_1	P_1	20	18	17
T_2	N_1	P_2	21	21	18
T_3	N_2	P_1	21	22	20
T_4	N_2	P_2	22	20	20
T_5	N_3	P_1	23	21	22
T_6	N_3	P_2	25	23	23
TESTIGO			15	17	15

2. García, G. (2016) realizó una investigación en la que evaluó la aplicación de un fertilizante foliar en tres dosis: 0.5, 0.75 y 1 ml/litro en tres épocas de aplicación. Estas épocas indican los momentos de aplicación a lo largo del ciclo del cultivo de tomate:

Época 1: Inicio floración + 15días después + 30 días después
Época 2: 15 DDT + Inicio floración + 45DDT (días después del trasplante)
Época 3: Inicio cosecha + 15DDC + 30DDC (días después de la cosecha).

El diseño utilizado fue bloques completos al azar con 3 repeticiones, bajo un arreglo en parcelas divididas, ubicando el factor **Época** en las parcelas grandes y **Dosis** en las parcelas pequeñas. Además se incluyó un tratamiento testigo (sin aplicación). La variable de respuesta medida fue el rendimiento de frutos de tomate expresados en kilogramos por hectárea. Los resultados se presentan en el cuadro siguiente.

			Bloques		
Tratamientos	Épocas	Dosis	I	II	III
T_1	E_1	D_1	29.09	39.73	37.64
T_2	E_1	D_2	36.09	40.36	41.55
T_3	E_1	D_3	37.09	45.09	30.82
T_4	E_2	D_1	42.45	26.00	32.91
T_5	E_2	D_2	48.29	34.45	31.09
T_6	E_2	D_3	37.82	40.55	44.00
T_7	E_3	D_1	41.36	44.64	42.73
T_8	E_3	D_2	29.73	41.82	31.27
T_9	E_3	D_3	47.18	33.82	34.91
TESTIGO			35.41	26.05	34.41

Realice el ANOVA y en caso necesario la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey.

3. Considere los resultados de un experimento reportado por Vaides, E. (2016), referentes a la evaluación de 3 fórmulas comerciales de fertilizante (triple quince, 12-20-12 y 10-30-10) y 3 dosis (50, 150 y 250 kg.ha^{-1}) más un testigo absoluto (sin aplicación), en una plantación de Teca (*Tectona grandis*), recién establecida en el departamento de Petén (Guatemala), la lectura fue a 6 meses después de plantados los árboles. La fertilización se hizo al momento de plantar, en dos posturas al lado de la planta sobre el camellón. La planta tenía una altura inicial de 20 cm con una diferencia de +- 2 cm. El diseño experimental utilizado fue el de bloques completos al azar, con 4 repeticiones. Las variables medidas fueron el diámetro (en cm a la altura del cuello de la planta) y la altura total (cm).

Realice el ANOVA y en caso necesario la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey. Los resultados se presentan en el cuadro siguiente:

Bloque	**Formula**	**Dosis**	**Tratamiento**	**Diametro**	**Altura**
1	15-15-15	50	T1	0.53	13.33
1	15-15-15	150	T2	0.56	15.20
1	15-15-15	250	T3	0.58	9.75
1	12-24-12	50	T4	0.63	19.33
1	12-24-12	150	T5	0.62	9.17
1	12-24-12	250	T6	0.66	14.00
1	10-30-10	50	T7	0.75	12.25
1	10-30-10	150	T8	0.87	16.33
1	10-30-10	250	T9	0.94	24.00
1	**Testigo**	**0**	T10	0.45	11.38
2	15-15-15	50	T1	0.70	15.33
2	15-15-15	150	T2	0.77	13.33
2	15-15-15	250	T3	0.48	16.40
2	12-24-12	50	T4	0.60	17.60
2	12-24-12	150	T5	0.60	12.00
2	12-24-12	250	T6	0.65	18.25
2	10-30-10	50	T7	0.53	8.67
2	10-30-10	150	T8	0.52	11.33
2	10-30-10	250	T9	0.50	16.38
2	**Testigo**	**0**	T10	0.45	11.17
3	15-15-15	50	T1	1.00	26.14
3	15-15-15	150	T2	0.68	14.17
3	15-15-15	250	T3	1.19	33.22
3	12-24-12	50	T4	0.56	14.80
3	12-24-12	150	T5	1.07	26.40

3	12-24-12	250	T6	1.27	34.90
3	10-30-10	50	T7	1.08	25.33
3	10-30-10	150	T8	1.40	29.70
3	10-30-10	250	T9	0.40	16.67
3	**Testigo**	**0**	T10	0.38	13.00
4	15-15-15	50	T1	1.01	23.81
4	15-15-15	150	T2	0.74	21.13
4	15-15-15	250	T3	1.63	35.90
4	12-24-12	50	T4	0.65	20.75
4	12-24-12	150	T5	2.17	51.89
4	12-24-12	250	T6	1.25	31.88
4	10-30-10	50	T7	0.62	18.17
4	10-30-10	150	T8	0.90	16.80
4	10-30-10	250	T9	0.79	18.75
4	**Testigo**	**0**	T10	0.58	16.25

4. En una investigación en maíz se determinó el efecto en la altura de dos fuentes de nitrógeno: urea y nitrato de amonio con tres dosis, 40, 80 y 120 $kg.ha^{-1}$. Se debe aclarar que el cálculo de las cantidades de nitrógeno, se hizo tomando en consideración las fuentes de este elemento. El ensayo fue realizado bajo un diseño de bloques completos al azar, con cinco repeticiones y un arreglo combinatorio A × B, en el que A corresponde a las fuentes y B a las dosis. Además se agregaron dos tratamientos, testigo 1, aplicación de materia orgánica y testigo 2, aplicación de bovinaza. Los resultados de altura (a los 90 días) expresados en centímetros se presentan a continuación:

			Bloques				
Tratamientos	Fuente	Dosis	I	II	III	IV	V
1	Urea	40	169.5	174.2	179.2	184.0	189.5
2	Urea	80	202.0	207.7	212.8	217.0	222.0
3	Urea	120	204.8	210.2	215.0	220.0	225.2
4	Nitrato_Amonio	40	163.0	167.7	172.8	177.5	183.6
5	Nitrato_Amonio	80	193.3	198.2	203.0	207.6	213.2
6	Nitrato_Amonio	120	198.9	203.8	209.5	215.0	224.0
Testigo 1			161.2	166.1	170.5	176.1	181.5
Testigo 2			180.1	185.5	179.0	187.8	175.6

Realice el ANOVA y en caso necesario la prueba de comparación múltiple de medias de acuerdo con el criterio propuesto por Tukey.

Referencias

ONOFRI, A. **2007**. Routine statistical analyses of field experiments by using an Excel extension. Proceedings 6th National Conference Italian Biometric Society: "La statistica nelle scienze della vita e dell'ambiente", Pisa, 20-22 June 2007, 93-96. Version 1.1 (Update: 18/03/2011).

CAPÍTULO 9
EXPERIMENTOS EN FRANJAS

9.1 INTRODUCCIÓN

Los experimentos en parcelas divididas (*split-plot*) son frecuentemente utilizados en la experimentación agronómica cuando, en un experimento de tipo factorial, el material experimental o la propia conducción del experimento no permite una completa aleatorización de todas las combinaciones de los niveles de los factores.

Una variación de los experimentos en parcelas divididas es caracterizada cuando, por diversas razones, aunque se estructuren las parcelas pequeñas, no hay posibilidad de una aleatorización total de estas parcelas. En este caso, las parcelas relativas a cada factor se posicionan en franjas (fajas, *strip plot* o *split -block*)), tanto en filas como en las columnas.

Considerando 2 factores, A y B, con a y b niveles, respectivamente, de forma que en cada nivel de A se tienen los niveles de B como subparcelas y viceversa, tanto los niveles de A como los niveles de B son considerados parcelas (factores principales), en tanto que la interacción (A×B) en cada uno de los índices de A y B constituyen las subparcelas (parcelas pequeñas).

Para ilustrar este caso, considere un experimento con dos factores A y B, con 3 y 4 niveles, respectivamente, y cada bloque constituido por fajas horizontales A_i y fajas verticales B_j, resultando, para un bloque cualquiera, el siguiente esquema:

	B_1 ↓	B_4 ↓	B_2 ↓	B_3 ↓
A_3 →	A_3B_1	A_3B_4	A_3B_2	A_3B_3
A_2 →	A_2B_1	A_2B_4	A_2B_2	A_2B_3
A_1 →	A_1B_1	A_1B_4	A_1B_2	A_1B_3

Y de igual manera para los demás bloques, los cuales difieren entre sí apenas en lo referente a la aleatorización.

De acuerdo con Gomes, FP (2000) el análisis de los experimentos en franjas son más complejos, en comparación con los dispuestos en un arreglo en parcelas divididas; y dan menor precisión a las comparaciones entre los niveles del factor que se ubica en la parcela pequeña. Por ello, los experimentos en franjas deben ser evitados, siempre que sea posible. Sin embargo, existen varias situaciones en que razones de orden práctica nos llevan a adoptarlos. Dentro de estas situaciones, se puede citar: experimentos con épocas de cosecha (cuando la cosecha es mecanizada), tipos de preparación del suelo, aplicación mecanizada de fertilizantes, aplicación de correcciones al suelo, aplicación de madurantes en caña de azúcar, en evaluaciones de distanciamientos de siembra, etc.

9.2 MODELO ESTADÍSTICO–MATEMÁTICO

Según Kempthorne (1952), citado por Nogueira (2007), el modelo adecuado con dos factores y un diseño bloques completos al azar es el siguiente:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \rho_k + (\rho\beta)_{jk} + (\alpha\rho)_{ik} + (\alpha\beta\rho)_{ijk} \quad \begin{cases} i = 1,2, \ldots, a \\ j = 1,2, \ldots, r \\ k = 1,2, \ldots, b \end{cases}$$

siendo:

Y_{ijk}	=	Variable de respuesta medida en la ijk - ésima unidad experimental
μ	=	Media general
β_j	=	Efecto del j - ésimo bloque
αi	=	Efecto del i - ésimo nivel del factor A.
$(\alpha\beta)_{ij}$	=	Efecto de la interacción entre el i-ésimo nivel del factor A con el j - ésimo bloque, o sea, es el error experimental asociado al factor A, tal que $(\alpha\beta)_{ij} \sim N(0, \sigma^2_1)$ e independientes, es utilizado como $error_{(a).}$
ρ_k	=	Efecto del k - ésimo nivel del factor B
$(\rho\beta)_{jk}$	=	Efecto de la interacción entre el k-ésimo nivel del factor A con el j - ésimo bloque, o sea, es el error experimental asociado al factor B, tal que $(\rho\beta)_{jk} \sim N(0, \sigma^2_2)$ e independientes, es utilizado como $error_{(b).}$
$(\alpha\rho)_{ik}$	=	Efecto debido a la interacción del i-ésimo nivel del factor A con el k - ésimo nivel del factor B.
$(\alpha\beta\rho)_{ijk}$	=	Error experimental asociado a Y_{ijk}, tal que $(\alpha\beta\rho)_{ijk} \sim N(0, \sigma^2)$ e independientes, es utilizado como término de error o residuo.

9.3 HIPÓTESIS

Ho: $\alpha_i = 0$, para todo i, contra, Ha: $\alpha_i \neq 0$, para algún i;
Ho: $\rho_k = 0$, para todo k, contra, Ha: $\rho_k \neq 0$, para algún k;
Ho: $(\rho\beta)_{ik} = 0$, para todo i y k, contra, Ha: $(\rho\beta)_{ik} \neq 0$, para algún i y k.

9.4 ESQUEMA DEL ANÁLISIS DE VARIANZA

FV	GL	SC	CM	Valor de F
Bloques	(r–1)	$SC_{Bloques}$		
A	(a–1)	SC_A	SC_A/GL_A	$CM_A/CM_{error\ (a)}$
Error (a)	(a–1) (r–1)	$SC_{error(a)}$	$SC_{error(a)}/GL_{error(a)}$	
B	(b–1)	SC_B	SC_B/GL_B	$CM_B/CM_{error\ (b)}$
Error (b)	(b–1) (r–1)	$SC_{error(b)}$	$SC_{error(b)}/GL_{error(b)}$	
A x B	(a–1) (b–1)	$SC_{A\ x\ B}$	SC_{AB}/GL_{AB}	$CM_{AxB}/CM_{residuo}$
Residuo	(a–1) (b–1) (r–1)	$SC_{residuo}$	$SC_{residuo}/GL_{residuo}$	
Total	abr–1	SC_{Total}		

Descripción	Suma de Cuadrados
Bloques	$SC_{BLOQUES} = \frac{\sum_{j=1}^{r} Y_{.j.}^{2}}{ab} - \frac{Y...^{2}}{abr}$
A	$SC_{A} = \frac{\sum_{i=1}^{a} Y_{i..}^{2}}{br} - \frac{Y...^{2}}{abr}$
A, Bloques	$SC_{(A,BLOQUES)} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{r} Y_{ij.}^{2}}{b} - \frac{Y...^{2}}{abr}$
error (a)	$SC_{(A\ x\ BLOQUES)} = SC_{(A,BLOQUES)} - SC_{A} - SC_{BLOQUES} = SC_{error(a)}$
B	$SC_{B} = \frac{\sum_{k=1}^{b} Y_{..k}^{2}}{ar} - \frac{Y...^{2}}{abr}$
B, Bloques	$SC_{(B,BLOQUES)} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{r} Y_{.jk}^{2}}{a} - \frac{Y...^{2}}{abr}$
error (b)	$SC_{(B\ x\ BLOQUES)} = SC_{(B,BLOQUES)} - SC_{B} - SC_{BLOQUES} = SC_{error(b)}$
A,B	$SC_{(A,B)} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{b} Y_{ik.}^{2}}{r} - \frac{Y...^{2}}{abr}$
A × B	$SC_{(A\ x\ B)} = SC_{(A,B)} - SC_{A} - SC_{B}$
Residuo	$SC_{residuo} = SC_{TOTAL} - SC_{BLOQUES} - SC_{A} - SC_{error(a)} - SC_{B} - SC_{error(b)} - SC_{AxB}$
Total	$SC_{TOTAL} = \sum_{i=1}^{a}\sum_{j=1}^{r}\sum_{k=1}^{b} Y_{ijk}^{2} - \frac{Y...^{2}}{abr}$

Decisión: Se rechaza Ho a un nivel α de significancia, cuando:

(1) $F_A \geq F(\alpha, a-1, (a-1)(r-1))$;
(2) $F_B \geq F(\alpha, b-1, (b-1)(r-1))$;
(3) $F_{AB} \geq F(\alpha, (a-1)(b-1), (a-1)(b-1)(r-1))$.

9.5 EJEMPLO DE APLICACIÓN

Los datos que se presentan a continuación se refieren al Pol% de la caña de azúcar, obtenido de un experimento en franjas en el ingenio Santa Cruz, citado por Nogueira (2007), que involucró 2 factores A y B; correspondientes a tipos de surcos, asociado con distanciamientos entre surcos y densidades de siembra, con 4 y 3 niveles respectivamente, siendo:

A_1: surco simple y distanciamiento de 1.40 m.
A_2: surco doble y distanciamiento de 1.40 m x 0.90 m.
A_3: surco de base larga y distanciamiento de 1.70 m.
A_4: surco de base larga y distanciamiento de 1.90 m.
B_1: 4 toneladas de semilla/ha
B_2: 6 toneladas de semilla/ha
B_3: 8 toneladas de semilla/ha

		Bloques				
Surcos	Densidades	I	II	III	IV	$Y_{ik.}$
A_1	B_1	17.67	17.23	17.43	17.61	69.94
	B_2	17.31	17.6	17.05	16.91	68.87
	B_3	17.49	17.3	17.68	18.27	70.74
		52.47	**52.13**	**52.16**	**52.79**	
A_2	B_1	17.19	17.85	17.44	17.56	70.04
	B_2	17.21	17.26	16.71	17.52	68.7
	B_3	18.04	16.38	17.23	17.14	68.79
		52.44	**51.49**	**51.38**	**52.22**	
A_3	B_1	17.39	17.54	16.61	17.51	69.05
	B_2	17.39	17.67	16.77	17.61	69.44
	B_3	17.69	17.02	17.34	18.02	70.07
		52.47	**52.23**	**50.72**	**53.14**	
A_4	B_1	17.19	17.57	17.72	17.73	70.21
	B_2	16.78	17.57	17.79	18.27	70.41
	B_3	17.86	16.85	18.12	17.94	70.77
		51.83	**51.99**	**53.63**	**53.94**	
	$Y_{.j.}$	209.21	207.84	207.89	212.09	837.03

Con estos datos se construyen los 3 cuadros auxiliares siguientes:

Cuadro Auxiliar 1

Surcos	Bloques I	II	III	IV	$Y_{i..}$
A_1	52.47	52.13	52.16	52.79	209.55
A_2	52.44	51.49	51.38	52.22	207.53
A_3	52.47	52.23	50.72	53.14	208.56
A_4	51.83	51.99	53.63	53.94	211.39
$Y_{.j.}$	209.21	207.84	207.89	212.09	837.03

Cuadro Auxiliar 2

Densidad	Bloques I	II	III	IV	$Y_{..k}$
B_1	69.44	70.19	69.2	70.41	279.24
B_2	68.69	70.1	68.32	70.31	277.42
B_3	71.08	67.55	70.37	71.37	280.37
$Y_{.j.}$	209.21	207.84	207.89	212.09	837.03

Cuadro Auxiliar 3

Surcos	Densidades B_1	B_2	B_3	$Y_{i..}$
A_1	69.94	68.87	70.74	209.55
A_2	70.04	68.7	68.79	207.53
A_3	69.05	69.44	70.07	208.56
A_4	70.21	70.41	70.77	211.39
$Y_{..k}$	279.24	277.42	280.37	837.03

Cálculo de las sumas de cuadrados

$$SC_{BLOQUES} = \frac{1}{12}(209.21^2 + \cdots + 212.09^2) - \frac{837.03^2}{48} = 0.9921$$

$$SC_A = \frac{1}{12}(209.55^2 + \cdots + 211.39^2) - \frac{837.03^2}{48} = 0.6753$$

$$SC_{(A,BLOQUES)} = \frac{1}{3}(52.47^2 + \cdots + 53.94^2) - \frac{837.03^2}{48} = 3.285$$

$$SC_{(A\ x\ BLOQUES)} = 3.285 - 0.6753 - 0.9921 = 1.6173 = SC_{error(a)}$$

$$SC_{B} = \frac{1}{16}(279.24^2 + \cdots + 280.37^2) - \frac{837.03^2}{48} = 0.2769$$

$$SC_{(B,BLOQUES)} = \frac{1}{4}(69.44^2 + \cdots + 71.37^2) - \frac{837.03^2}{48} = 3.5625$$

$$SC_{(B\ x\ BLOQUES)} = 3.5625 - 0.2769 - 0.9921 = 2.2935 = SC_{error(b)}$$

$$SC_{(A,B)} = \frac{1}{4}(69.94^2 + \cdots + 70.77^2) - \frac{837.03^2}{48} = 1.5687$$

$$SC_{(A\ x\ B)} = 1.5687 - 0.6753 - 0.2769 = 0.6165$$

$$SC_{TOTAL} = (17.67^2 + \cdots + 17.94^2) - \frac{837.03^2}{48} = 8.4247$$

$$SC_{RESIDUO} = 8.4247 - (0.9921 + 0.6753 + 1.6173 + 0.2769 + 2.2935 + 0.6165) = 1.9531$$

Cuadro del análisis de varianza

FV	GL	SC	CM	Valor de F	F crítica
Bloques	3	0.9921	0.3307		
A	3	0.6753	0.2251	1.253	3.86
Error (a)	9	1.6173	0.1797		
B	2	0.2769	0.1385	0.362	5.14
Error (b)	6	2.2935	0.3823		
A × B	6	0.6165	0.1027	0.947	2.66
Residuo	18	1.9531	0.1085		
Total	47	8.4247			

Conclusión:

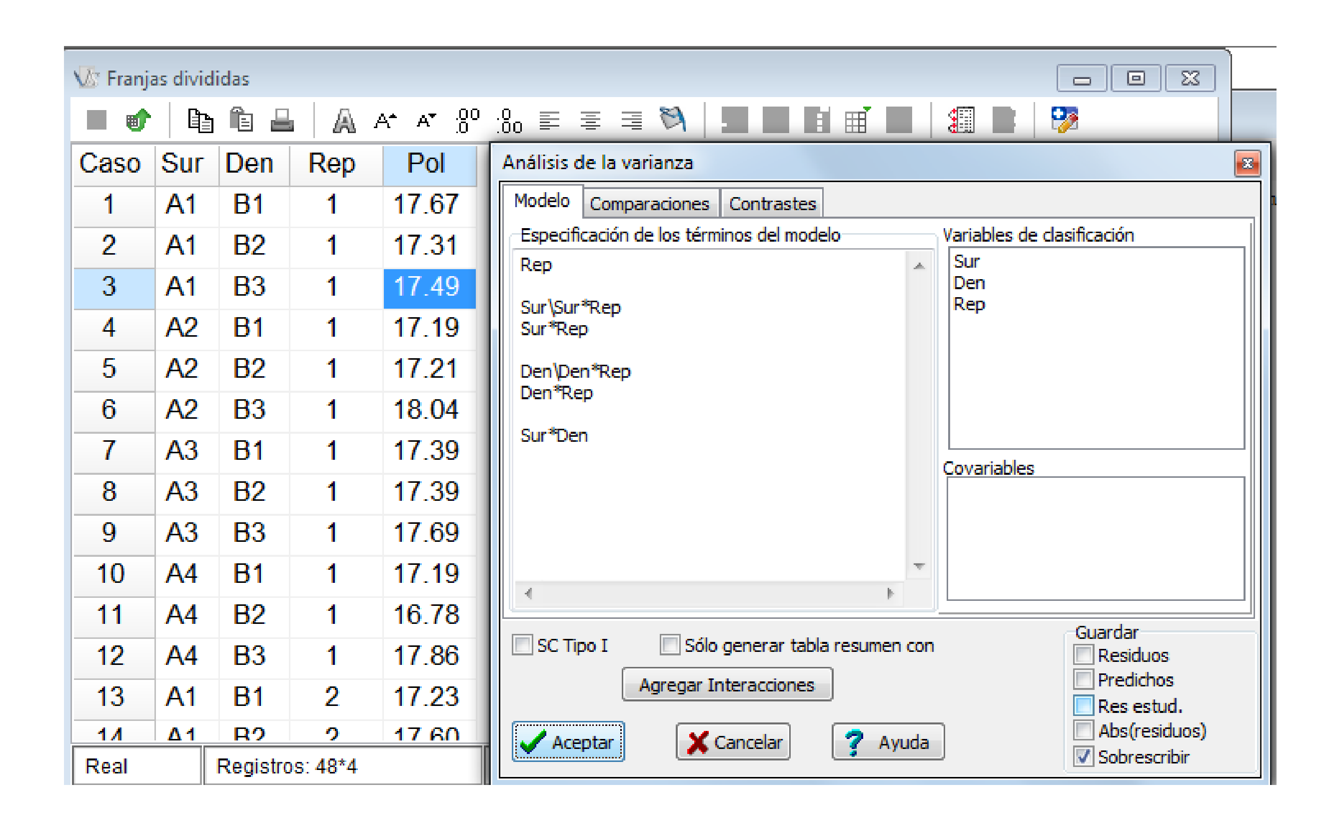
Por medio de la prueba de F no fue constatado efectos de los tipos de surcos, asociado con distanciamientos entre surcos, ni de las densidades de siembra, y tampoco con la interacción de los niveles de estos dos factores.

9.6 PROGRAMA EN SAS PARA UN EXPERIMENTO BIFACTORIAL CON ARREGLO EN FRANJAS DIVIDIDAS

```
OPTIONS nodate nonumber;
TITLE "Experimento en franjas";
DATA fran;
        DO surcos=1 TO 4;
        DO densid=1 TO 3;
        DO bloques=1 TO 4;
INPUT pol @@; OUTPUT;
        END;
        END;
        END;
CARDS;
17.67   17.23   17.43   17.61
17.31   17.6    17.05   16.91
17.49   17.3    17.68   18.27
17.19   17.85   17.44   17.56
17.21   17.26   16.71   17.52
18.04   16.38   17.23   17.14
17.39   17.54   16.61   17.51
17.39   17.67   16.77   17.61
17.69   17.02   17.34   18.02
17.19   17.57   17.72   17.73
16.78   17.57   17.79   18.27
17.86   16.85   18.12   17.94
;
PROC print; RUN;
TITLE "Análisis de varianza y prueba de Tukey";
PROC glm;
        CLASS surcos densid bloques;
        MODEL pol = bloques surcos bloques*surcos densid bloques*densid surcos*densid;
        TEST h=surcos e=bloques*surcos;
        TEST h=densid e=bloques*densid;
MEANS surcos/TUKEY e=bloques*surcos;
MEANS densid/TUKEY e=bloques*densid;
MEANS surcos*densid;
RUN;
```

Para resolver el ejemplo en Infostat, se debe de considerar los errores de cada factor e indicarlos, pues es necesario para que al realizar la comparación de medias se utilice el error adecuado.

Recuerde organizar los datos por columnas. Estas columnas las hemos etiquetado como: Sur =tipo de surco, Den = densidad de siembra, Rep = bloque o repetición y Pol = Pol%. Luego, en la especificación del modelo, digite lo siguiente:



La salida que proporciona Infostat se muestra a continuación:

Análisis de la varianza

Variable	N	R²	R² Aj	CV
Pol	48	0.77	0.39	1.89

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor	(Error)
Rep	0.99	3	0.33	3.05	0.0554	
Sur	0.68	3	0.23	1.25	0.3473	(Sur*Rep)
Sur*Rep	1.62	9	0.18	1.66	0.1731	
Den	0.28	2	0.14	0.36	0.7104	(Den*Rep)
Den*Rep	2.29	6	0.38	3.52	0.0175	
Sur*Den	0.62	6	0.10	0.95	0.4869	
Error	1.95	18	0.11			
Total	8.42	47				

Compare los resultados obtenidos con el procedimiento manual. Discuta los resultados y redacte las conclusiones.

9.7 EJERCICIOS PROPUESTOS

1. Los siguientes datos se refieren a la producción de trigo (*Triticum* spp.) obtenido en un experimento donde se utilizó un diseño bloques al azar con arreglo en franjas divididas, con el objetivo de estudiar el efecto de dos sistemas de riego (R_1 y R_2) y tres dosis de nitrógeno (N_0, N_1 y N_2), distribuidos en cuatro bloques.

		Bloques			
Dosis de N	**Sistema de riego**	**I**	**II**	**III**	**IV**
N_0	R_1	55	63	63	65
N_0	R_2	71	77	77	75
N_1	R_1	62	66	70	66
N_1	R_2	77	79	78	76
N_2	R_1	69	77	79	76
N_2	R_2	78	81	80	79

Realice el ANOVA, interprete los resultados y concluya.

2. Estrada, RA (1993) realizó el trabajo de tesis titulado: "Evaluación del efecto de 16 distancias de siembra sobre el crecimiento y rendimiento en el cultivo de chile chocolate *(Capsicum* sp) en el Valle Central de Guatemala. Los factores y niveles sometidos a evaluación fueron los siguientes:

	Niveles			
Factor	**1**	**2**	**3**	**4**
A. Distancia entre surcos (m)	0.25	0.50	0.75	1.0
B. Distancia entre plantas (m)	0.25	0.50	0.75	1.0

El diseño utilizado fue bloques al azar, con un arreglo en franjas y tres repeticiones. La unidad experimental estuvo formada por 30 plantas, delimitándose una unidad de muestreo de 12 plantas. Se trazaron 6 surcos con 5 plantas dentro de cada surco. El largo de cada bloque fue de 10 m. y el ancho de 12.5 m, dejando una separación de 1.5 m entre bloques. Con esta información responda las siguientes interrogantes:

a) Describa el modelo estadístico–matemático asociado a este experimento.

b) Describa las combinaciones a analizar.

c) A su criterio, ¿qué variables de respuesta mediría?

d) Dibuje un croquis del experimento, mostrando la distribución aleatorizada de los tratamientos.

3. A continuación se presenta el croquis de campo de un experimento con remolacha (*Beta vulgaris*), realizado en un diseño de bloques al azar con 4 repeticiones y arreglo en franjas. Fueron evaluados dos factores: Fertilización nitrogenada (N_0 = 0, N_1 = 80, N_2 = 160 y N_3 = 320 libras/acre) y época de siembra (E_1, E_2, E_3, E_4 y E_5). La variable de respuesta medida fue la producción de remolacha en toneladas por acre.

		N_1	N_2	N_0	N_3
	E_4	26.4	34.2	24.8	30.2
	E_5	29.3	30.3	29.2	30.8
Bloque 1	E_1	10.1	10.8	8.4	10.4
	E_3	23.1	22.4	20.7	24
	E_2	18.2	18.5	15.6	22.4

		N_3	N_0	N_1	N_2
	E_4	31.2	21.3	26	29.2
	E_2	19.2	12.5	16.9	20.9
Bloque 2	E_3	25.9	16.7	21.2	24.3
	E_5	34.2	19.1	31	35.2
	E_1	10.3	5.2	10.8	11.2

		N_1	N_2	N_0	N_3
	E_1	31.2	21.3	26	29.2
	E_5	19.2	12.5	16.9	20.9
Bloque 3	E_2	25.9	16.7	21.2	24.3
	E_3	34.2	19.1	31	35.2
	E_4	10.3	5.2	10.8	11.2

		N_0	N_3	N_2	N_1
	E_4	10.1	31.9	28.7	23.1
	E_3	9.8	22.8	22.8	20.9
Bloque 4	E_5	11.4	29.2	32.6	23.2
	E_1	2.3	7.4	8.5	9
	E_2	8.8	17.8	17.2	15.9

Fuente: Banzatto y Kronka (2011)

a) Realice el ANOVA, interprete los resultados y concluya.
b) En caso de ser necesario aplique un análisis postANOVA.

CAPÍTULO 10
ANÁLISIS DE COVARIANZA (ANCOVA)

10.1 INTRODUCCIÓN

Uno de los problemas con el que frecuentemente se enfrenta el investigador, es el de controlar aquellos factores que no le he es posible medir y cuyo efecto no puede justificar, los cuales constituyen el error experimental. Una de las formas de minimizar este error es mediante la aleatorización de los tratamientos y la utilización de material experimental muy homogéneo. Sin embargo, la aleatorización difícilmente cancela la influencia de las variables involucradas en el error y la disponibilidad de material experimental homogéneo no es frecuente en algunos experimentos, principalmente con animales, quedando restringidos a experimentos de laboratorio, invernadero o con animales de bioterio.

Ronald Fisher en 1932 desarrolló una técnica conocida como **Análisis de Covarianza**, que combina el Análisis de Regresión con el Análisis de Varianza. **Covarianza** significa variación simultánea de dos variables que se asume están influyendo sobre la variable respuesta. En este caso se tiene la variable independiente tratamientos y otra variable que no es efecto de tratamientos pero que influye en la variable de respuesta, llamada a menudo: **covariable**.

El Análisis de Covarianza consiste básicamente en elegir una o más variables adicionales o covariables que estén relacionadas con la variable de respuesta, evitando que los promedios de tratamientos se confundan con los de las covariables, incrementando de esa manera la precisión del experimento. Por ejemplo: número de plantas por unidad experimental, pesos iniciales en animales, grado de infestación de garrapatas, días de lactancia o edad de destete, etc.; pueden ser covariables que influyan en el resultado final y cuyo efecto de regresión sobre la variable respuesta el investigador desea eliminar, ajustando las medias de tratamientos a una media común de X. En este análisis se asume que la variable dependiente Y está asociada en forma lineal con la variable independiente X, existiendo homogeneidad de pendientes. Como regla general para decidir sobre el empleo de la covarianza, el investigador debiera tener la certeza de que sus covariables no están influenciadas por los tratamientos estudiados.

El procedimiento de análisis comprende:

a) ANDEVA para X (covariable),

b) ANDEVA para Y (variable de respuesta),

c) Estimación del coeficiente angular de la regresión.

d) Obtención de la ecuación de regresión y ajuste a los promedios de la variable de respuesta.

10.2 SUPOSICIONES BÁSICAS DEL ANÁLISIS DE COVARIANZA

Como es de esperarse, las suposiciones que se hacen cuando se efectúa un análisis de covarianza son similares a las requeridas para la regresión lineal y el análisis de varianza. De esta manera, se encuentran las suposiciones usuales de independencia, normalidad, homocedasticidad, X fijas, etc. Para ser más exactos, se presenta a continuación los modelos estadístico−matemáticos asociados con algunos de los diseños más comunes cuando se realiza un análisis de covarianza.

a) Diseño Completamente al Azar

$$Y_{ij} = \mu + \tau_i + \beta (X_{ij} - \bar{X}_{..}) + \varepsilon_{ij} \qquad \begin{cases} i = 1, \ldots, t \\ j = 1, \ldots, r \end{cases}$$

Y_{ij} = Variable de respuesta medida en la j−ésima repetición y el i−ésimo tratamiento.
μ = Media general
τ_i = Efecto del i−ésimo tratamiento.
β = Coeficiente angular de la regresión.
X_{ij} = Variable independiente o covariable.
$\bar{X}_{..}$ = Media general de la covariable.
ε_{ij} = Error experimental.

b) Diseño en bloques completos al azar.

$$Y_{ij} = \mu + \tau_i + \rho_j + \beta (X_{ij} - \bar{X}_{..}) + \varepsilon_{ij} \qquad \begin{cases} i = 1, \ldots, t \\ j = 1, \ldots, r \end{cases}$$

Y_{ij} = Variable de respuesta medida en la j−ésima repetición y el i−ésimo tratamiento.
μ = Media general
τ_i = Efecto del i−ésimo tratamiento.
ρ_j = Efecto del j−ésimo bloque o repetición.
β = Coeficiente angular de la regresión.
X_{ij} = Variable independiente o covariable.
$\bar{X}_{..}$ = Media general de la covariable.
ε_{ij} = Error experimental.

c) Diseño cuadrado latino

$$Y_{ijk} = \mu + \tau_i + \rho_j + \gamma_k + \beta (X_{ijk} - \bar{X}_{..}) + \varepsilon_{ijk} \qquad \begin{cases} i = 1, \ldots, t \\ k = 1, \ldots, t \\ j = 1, \ldots, r \end{cases}$$

Y_{ijk} = Variable de respuesta medida en la j−ésima repetición y el i−ésimo tratamiento.
μ = Media general
τ_i = Efecto del i−ésimo tratamiento.
ρ_j = Efecto de la j−ésima fila.
γ_k = Efecto de la k−ésima columna.
β = Coeficiente angular de la regresión.
X_{ijk} = Variable independiente o covariable.
$\bar{X}_{..}$ = Media general de la covariable.
ε_{ijk} = Error experimental.

Otra suposición necesaria para el análisis correcto de covarianza, es que la variable concomitante X, no debe ser afectada por los tratamientos.

10.3 EJEMPLO DE APLICACIÓN

Un grupo de estudiantes del curso de Investigación Agrícola de la Escuela Nacional Central de Agricultura evaluó en 1990 el efecto del tiempo de cosecha sobre el rendimiento de grano de maíz. Se utilizaron 4 tratamientos y 3 repeticiones, con el diseño bloques completos al azar. Los tratamientos fueron: 30, 40, 50 y 60 días después de la polinización. El número de plantas (x) planificado por parcela útil fue de 52, pero al cosechar se obtuvieron diferentes números de plantas por unidad experimental. Los valores de la variable respuesta (y) se presentan en el cuadro siguiente:

	Repeticiones							
Tratamientos	I		II		III			
	x	y	X	Y	x	y	$x_{i.}$	$y_{i.}$
30	41	4.08	24	2.78	31	2.79	96	9.65
40	37	4.72	32	4.92	38	4.50	107	14.14
50	37	4.00	34	5.05	47	5.54	118	14.59
60	35	4.59	22	3.63	44	6.20	101	14.42
$x_{.j}$, $y_{.j}$	150	17.39	112	16.38	160	19.03	**422**	**52.80**

1º. ANDEVA para las variables X y Y

ANDEVA para X

FV	GL	SC	CM	Valor F	F crítica
Trats.	3	89.67	29.89	0.98	4.76
Bloques	2	320.67	160.33		
Error	6	183.33	30.56		
Total	11	593.67			

ANDEVA para Y

FV	GL	SC	CM	Valor F	F crítica
Trats.	3	5.64	1.88	2.29	4.76
Bloques	2	0.89	0.45		
Error	6	4.92	0.82		
Total	11	11.45			

X = número de plantas

Y = rendimiento expresado en kilogramos/unidad experimental.

2º. Estimación del coeficiente angular de la regresión y del coeficiente de correlación.

Factor de corrección = F.C. (*x*,*y*) = $$\frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{r} x_{ij}\right)\left(\sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}\right)}{tr} = \frac{(422)\,(52.8)}{4\times 3} = 1856.8$$

Suma de Productos Total para *x* y *y*. SPT(*x*,*y*)

$$SPT(x,y) = \sum_{i=1}^{t}\sum_{j=1}^{r} x_{ij}\, y_{ij} - F.C.(x,y) = [(41\times 4.08) + (37\times 4.72) + ... + (44\times 6.2)] - 1856.8 = 60.16$$

Suma de Productos de Bloques para *x* y *y*. SPB (*x*,*y*)

$$SPB(x,y) = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r} x_{.j}\, y_{.j}}{t} - F.C.(x,y) = \frac{[(150\times 17.39) + ... + (160\times 19.03)]}{4} - 1856.8 = 15.165$$

Suma de Productos de Tratamietos para *x* y *y*. SPTrat (*x,y*)

$$SPTrat(x,y) = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r} x_{.j}\, y_{.j}}{r} - F.C.(x,y) = \frac{[(96\times 9.65)+...+(101\times 14.42)]}{3} - 1856.8 = 15.673$$

Suma de Productos del Error para *x* y *y*. SPE (*x,y*)

$$SPE(x,y) = SPT(x,y) - [\, SPTrat(x,y) + SPB(x,y)\,] = 60.16 - (15.673 + 15.165) = 29.322$$

Coeficiente angular de la regresión: $\hat{\beta} = \frac{SPE(x,y)}{SCE(x)} = \frac{29.322}{183.33} = 0.1599 \approx \mathbf{0.16}$

Este coeficiente da la relación promedio de rendimiento por planta, es decir, el efecto de una planta en promedio es de 0.1599 kg.

Debe aclararse que el coeficiente de regresión β se supuso diferente de cero. Si este no fuera el caso, la introducción de la variable concomitante X sería una complicación innecesaria. Algunas veces el investigador querrá comprobar estas suposiciones. Esto es, evaluará las hipótesis:

Ho: $\beta = 0$ (no hay regresión lineal simple)
Ha: $\beta \neq 0$

Utilizando la estadística F (Fisher–Snedecor)

$$F = \frac{[SPE(x,y)]^2 / SCE(x)}{CME(y\ ajustado)} = \frac{29.322^2 / 183.33}{0.04666318} = 100.50$$

que tiene $v_1 = 1$ y $v_2 = (r-1)(t-1) - 1$, grados de libertad. En este caso $F_{crítica}(1,5,0.05) = 6.61$. Por lo tanto se concluye que la regresión lineal es significativa. El valor del CME (y ajustado) consúltelo en la Tabla resumen del ANCOVA que se presenta en la siguiente página.

El cálculo del coeficiente de correlación lineal (r) se efectúa de la manera siguiente:

$$r = \frac{SPE(x,y)}{\sqrt{SCE(x)\times SCE(y)}} = \frac{29.322}{\sqrt{(183.33)\times(4.92)}} = 0.976$$

Este valor de r puede ser evaluado con la prueba t de Student:

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n} = \frac{0.976}{\sqrt{1-(0.976)^2}} \times \sqrt{5} = 10.02^*$$

t crítica (5,0.05/2) = 2.57

siendo n = 5, el número de grados de libertad del residuo, luego de ser ajustado por la regresión (se le restó un grado de libertad).

3º. ANCOVA (los valores ajustados)

a) Cálculo de la suma de cuadrados de la regresión lineal.

$$SC\,Reg = \frac{[SPE(x,y)]^2}{SCE(x)} = \frac{(29.322)^2}{183.33} = 4.69$$

b) Suma de cuadrados del residuo, ajustada a la regresión

$$SCE(y\ Ajustado) = SCE(y) - SC\,Reg = 4.92 - 4.69 = 0.23$$

c) Suma de cuadrados de los tratamientos, ajustada de acuerdo con la regresión

$$SCTrat\ Ajustada = [SCE(y) + SCTrat(y)] - \frac{[\,SPTrat(x,y) + SPE(x,y)\,]^2}{SCTrat(x) + SCE(x)} - SCE(y\ Ajustada)$$

$$SCTrat\ Ajustada = [4.92 + 5.64] - \frac{(15.673 + 29.332)^2}{(89.67 + 183.33)} - 0.23 = 2.914$$

d) Resumen del ANCOVA

FV	GL	SCX	SCY	Suma de Productos
Tratamientos	3	89.67	5.64	15.673
Bloques	2	320.67	0.89	15.165
Error	6	183.33	4.92	29.322
Total	11	593.67	11.45	60.16

GL	SC	CM	Valor F	F crítica
3	2.914	0.97	21.09*	5.41
5	0.23	0.04666		
10				

De acuerdo con el ANCOVA, existen diferencias significativas entre tratamientos. En consecuencia, es conveniente hacer un ajuste por número de plantas a los promedios de rendimiento, de acuerdo con la siguiente ecuación:

$$\hat{y}_{i.} = \bar{y}_{i.} - \hat{\beta}\,(\bar{x}_{i.} - \bar{x})\ ,\ \text{siendo:}$$

$\hat{y}_{i.}$ = promedio ajustado de cada tratamiento.

$\bar{y}_{i.}$ = promedio de cada tratamiento sin ajustar.

$\hat{\beta}$ = coeficiente angular de la regresión.

$\bar{x}_{i.}$ = promedio del número de plantas de cada tratamiento.

$\bar{x}$ = promedio general del número de plantas.

e) El error estándar para la diferencia SE(d) entre dos medias ajustadas es dado por:

$$SE(d) = \sqrt{CME(\text{y Ajustado}) \times \left[\frac{1}{r_i} + \frac{1}{r_j} + \frac{(\overline{x}_i - \overline{x}_j)^2}{SCE(x)} \right]}$$

Cuando el número de repeticiones es el mismo para todos los tratamientos, el error estándar para la diferencia entre dos medias ajustadas es dado por:

$$SE(d) = \sqrt{\frac{2 \times CME(\text{y Ajustado})}{r} \times \left[1 + \frac{(\overline{x}_i - \overline{x}_j)^2}{SCE(x)} \right]}$$

Cuando los valores de $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_t$ no son muy diferentes (lo que se puede concluir cuando los tratamientos no producen efectos significativos en la variable X), se puede usar una estimación media para el error estándar, aplicable a cualquier contraste entre dos tratamientos. Esta estimación media tiene la siguiente expresión:

$$SE(d) = \sqrt{\frac{2 \times CME(\text{y Ajustado})}{r} \times \left[1 + \frac{CMTrat(x)}{SCE(x)} \right]}$$

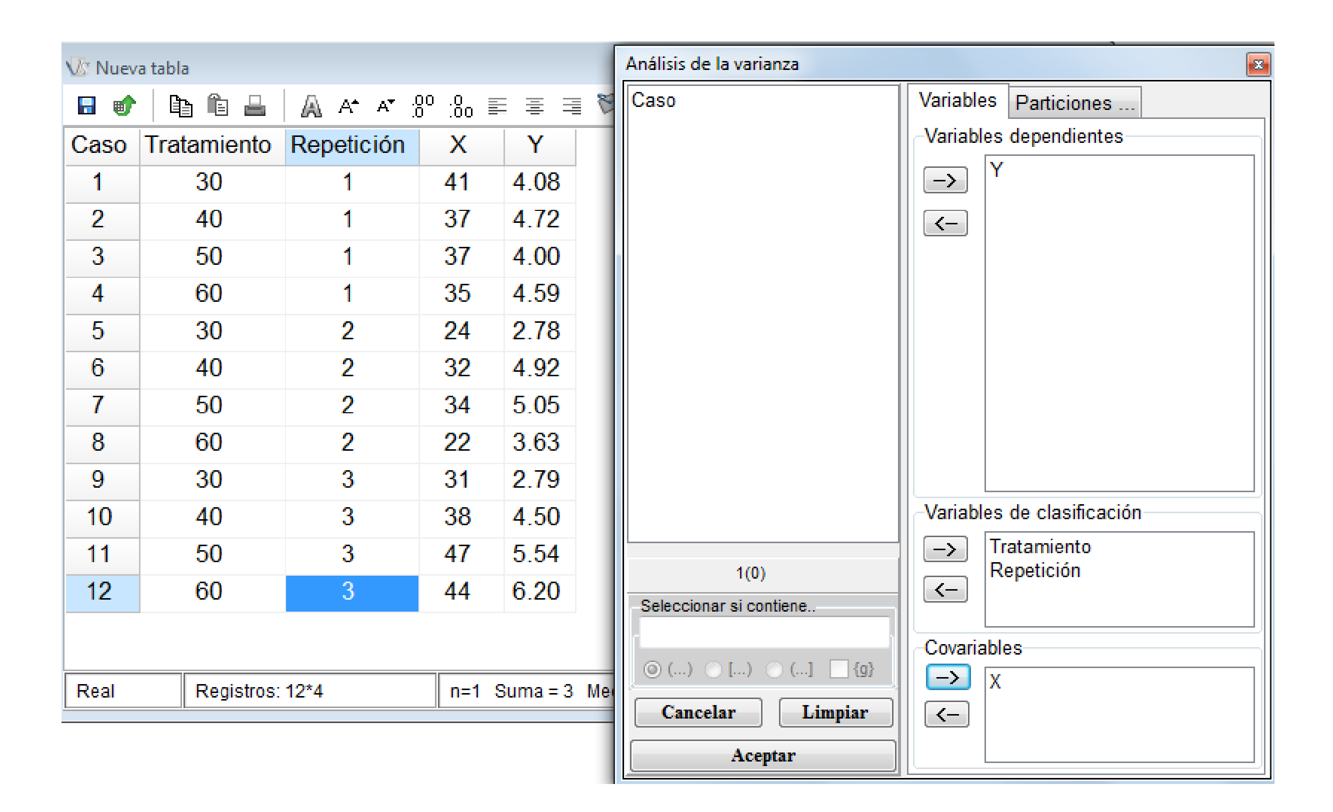
f) El cálculo de los promedios ajustados del rendimiento de grano de maíz $\hat{y}_{i.}$ se presenta a continuación:

Tratamientos	$y_{i.}$	$\overline{y}_{i.}$	$x_{i.}$	$\overline{x}_{i.}$	$\hat{y}_{i.}$
30	9.65	3.22	96	32.00	3.72
40	14.14	4.71	107	35.67	4.63
50	14.59	4.86	118	39.33	4.20
60	14.42	4.81	101	33.67	5.05
			Media general	**35.17**	

g) La presentación final de los resultados quedará de la siguiente forma:

Días después de la polinización	Media ajustada (kg/u.exp.)	
60	5.05	a
40	4.63	a b
50	4.20	b c
30	3.72	c

h) Compare los resultados anteriores, con los obtenidos en Infostat. Para ello ingrese los datos de la siguiente forma e informe las variables como se muestra en el cuadro de Análisis de Varianza.



Los resultados generados por Infostat se muestran a continuación:

Análisis de la varianza

Variable	N	R^2	R^2 Aj	CV
Y	12	0.98	0.96	4.91

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor	Coef
Tratamiento	2.91	3	0.97	20.78	0.0030	
Repetición	1.66	2	0.83			
X	4.69	1	4.69	100.50	0.0002	0.16
Error	0.23	5	0.05			
Total	11.45	11				

Test:Tukey Alfa=0.05 DMS=0.65082

Error: 0.0467 gl: 5

Tratamiento	Medias	n	E.E.			
60	5.05	3	0.13	A		
40	4.63	3	0.12	A	B	
50	4.20	3	0.14		B	C
30	3.72	3	0.13			C

Medias con una letra común no son significativamente diferentes (p<= 0.05)

10.4 PROGRAMA EN SAS PARA UN EXPERIMENTO DONDE SE UTILIZÓ ANÁLISIS DE COVARIANZA SIMPLE

```
OPTIONS nodate nonumber;
DATA ancova;
TITLE “Analisis de Covarianza Simple”;
INPUT trat rep x y;
CARDS;
30  1  41  4.08
40  1  37  4.72
50  1  37  4.00
60  1  35  4.59
30  2  24  2.78
40  2  32  4.92
50  2  34  5.05
60  2  22  3.63
30  3  31  2.79
40  3  38  4.50
50  3  47  5.54
60  3  44  6.20
;
PROC glm;
CLASS trat rep;
MODEL y=trat rep/SS1;
PROC glm;
CLASS trat rep;
MODEL y=trat rep x;
MEANS trat;
LSMEANS trat/stderr pdiff;
RUN;
```

10.5 EJERCICIOS PROPUESTOS

1. Un ensayo fue realizado para evaluar el control de plagas en el cultivo del frijol. En ese experimento se utilizó un diseño en bloques al azar y se evaluaron 5 tratamientos: Testigo, Disyston, Ekatin, Keltane y Diazinon. Además de la producción (Y), expresada en gramos por unidad experimental, se contó el número de plantas de cada parcela (X). Los datos se presentan en la tabla siguiente:

		Tratamientos				
Bloque	**Variables**	**Testigo**	**Disyston**	**Ekatin**	**Keltane**	**Diazinon**
I	X	9	7	9	6	8
	Y	74	58	118	41	95
II	X	9	8	9	9	8
	Y	51	67	48	38	41
III	X	8	5	9	8	9
	Y	95	40	49	77	39
IV	X	9	8	9	9	9
	Y	62	58	64	92	114
V	X	9	6	8	7	6
	Y	60	29	67	57	35
VI	X	9	8	8	7	8
	Y	47	64	51	77	49
VII	X	6	9	8	8	9
	Y	14	55	15	59	39
VIII	X	8	8	9	8	9
	Y	19	47	29	32	100

Con estos datos realice el Análisis de Covariaza y concluya.

2. Con el objetivo de evaluar 4 tipos de raciones en cerdos, se realizó una prueba de engorda. El diseño experimental utilizado fue completamente al azar con 6 repeticiones. Los datos de ganancia en peso (Y) y pesos iniciales (X) de los cerdos, se presentan en el cuadro siguiente:

Tratamientos							
1		2		3		4	
X	Y	X	Y	X	Y	X	Y
30	165	24	180	34	156	41	201
27	170	31	169	32	189	32	173
20	130	20	171	35	138	30	200
21	156	26	161	35	190	35	193
33	167	20	180	30	160	28	142
29	151	25	170	29	172	36	189

Con estos datos realice el Análisis de Covariaza y concluya.

3. La siguiente tabla muestra la altura inicial (x) y la altura alcanzada luego de cuatro meses (y) de plantas de *Leucaena leucocephala*, en tres tipos diferentes de suelo en un área experimental. Se utilizó un diseño completamente al azar, con diez repeticiones. Ambas alturas son expresadas en centímetros.

Parcela	Tratamiento 1		Tratamiento 2		Tratamiento 3	
	X	Y	X	Y	X	Y
1	18	145	27	161	31	180
2	22	149	28	164	27	158
3	26	156	27	172	34	183
4	19	151	25	160	32	175
5	15	143	21	166	35	195
6	25	152	30	175	36	196
7	16	144	21	156	35	187
8	28	154	30	175	23	137
9	23	150	22	158	34	184
10	24	151	25	165	32	184

Con estos datos realice el Análisis de Covariaza y concluya.

4. Un Ingeniero Agroindustrial está estudiando tres técnicas diferentes de deshidratación con el propósito de industrializar una fruta. Utiliza un diseño completamente al azar para evaluar el índice de recuperación del agua en cada fruta tratada. Debido a que el tamaño del fruto estudiado varía, se pesa cada uno de ellos al asignarlo al tratamiento. En este caso el peso es una covariable en el proceso. Así, el factor es la técnica (t) de deshidratación, la respuesta y: el índice de rehidratación y la covariable x; el peso en gramos. Los datos que resultaron al realizar el experimento se presentan a continuación:

Tratamiento	y	x	Tratamiento	y	x	Tratamiento	y	x
1	57	11.5	2	77	15.5	3	58	14.5
1	60	13.0	2	89	16.5	3	64	15.0
1	69	15.0	2	90	18.0	3	73	18.0
1	71	14.0	2	92	19.5	3	75	17.5
1	81	17.0	2	104	23.0	3	78	19.0
1	83	18.5	2	101	22.5	3	80	20.0

Fuente: Castaño, E y Domínguez, J. (2001)

Con estos datos, realice el Análisis de Covarianza y emita las conclusiones respectivas.

5. En un cultivo de sorgo forrajero se evaluó el rendimiento de grano y la producción de forraje. Para ellos, se compararon 7 tratamientos: 30, 35, 40, 45, 50, 55 y 60 días después de ocurrida la floración. El número de plantas por parcela útil fue de 52. A la cosecha el número de plantas por tratamiento y repetición fue diferente. Los datos obtenidos al finalizar el experimento se presentan en el siguiente cuadro, en el cual X representa el número de plantas por parcela y Y la producción de grano en kilogramos por parcela. El diseño experimental utilizado fue el de bloques completos al azar con 6 repeticiones.

	Bloques											
	I		II		III		IV		V		VI	
Días	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
30	41	4.08	24	2.78	31	2.79	46	4.24	32	4.17	38	2.62
35	40	4.26	36	4.23	44	5.6	48	6.36	47	4.33	47	4.03
40	37	4.72	32	4.92	38	4.5	41	5.62	42	5.15	40	4.32
45	32	4.25	38	4.53	40	4.83	40	4.3	51	5.43	46	4.52
50	37	4	34	5.05	47	5.54	41	6.46	50	6.65	39	5.7
55	42	6.16	38	5	34	4.61	40	5.41	42	5.13	49	5.86
60	35	4.59	22	3.63	44	6.2	39	5.47	40	5.16	40	6.07

Con estos datos realice el Análisis de Covariaza y concluya.

6. Teos Morales (1980) realizó el trabajo de tesis titulado "Determinación del nivel de tolerancia de la planta de maíz (*Zea mays* L.) al daño causado por el gusano cogollero (larva de *Spodoptera frugiperda* J. E. Smith)". El experimento fue instalado en la parcela 358, Línea B4, del Parcelamiento Agrário "San José La Máquina". Fue utilizada semilla de maíz híbrido H-5, originario de El Salvador.

   El diseño experimental utilizado fue Cuadrado Latino, siendo el área total utilizada: 1653.55 m^2 (87.5 m × 18.90 m). El área bruta de cada unidad experimental fue de 33.75 m^2 (12.5 m × 2.70 m) y el área neta de 21.60 m^2, constituida por 2 surcos de 12 m de longitud. Entre las variables de respuesta analizadas, está el rendimiento de grano de maíz, expresado en libras por unidad experimental. También se realizó el conteo de plantas por parcela (utilizado como covariable). Los tratamientos evaluados en la investigación, se refieren a las diferentes áreas foliares consumidas por el cogollero; para definir esas áreas, se realizó diferente número de aplicaciones del insecticida volatón granulado en dosis de l5 libras por manzana, directamente al cogollo de la planta, a diferente altura y fecha, como se indica en el cuadro siguiente (en el interior del cuadro se muestra el número de aplicaciones del insecticida):

	Altura de planta (m)					
Tratamiento	0.05	0.15	0.25	0.40	0.60	1.00
A	1	2	3	4	5	6
B		1	2	3	4	5
C			1	2	3	4
D				1	2	3
E					1	2
F						1
G						0

La cosecha fue realizada de forma manual a los 140 días después de la siembra. Los resultados se presentan en el croquis de campo. En negrito el número de plantas por unidad experimental.

	Columna						
Fila	1	2	3	4	5	6	7
1	F 22.250 **85**	B 17.000 **82**	D 24.125 **84**	G 24.250 **80**	E 22.738 **83**	C 21.500 **79**	A 22.500 **80**
2	C 23.500 **90**	F 22.500 **84**	A 26.500 **82**	D 27.000 **80**	B 27.000 **75**	G 20.875 **77**	E 26.500 **82**
3	G 23.500 **85**	C 22.000 **81**	E 26.000 **86**	A 28.875 **89**	F 23.875 **73**	D 23.000 **77**	B 24.000 **79**
4	A 25.875 **80**	D 24.312 **95**	F 19.875 **75**	B 24.250 **76**	G 25.250 **85**	E 25.250 **85**	C 20.500 **77**
5	D 22.312 **85**	G 15.500 **76**	B 22.875 **74**	E 20.500 **77**	C 22.500 **82**	A 28.000 **89**	F 22.250 **79**
6	E 17.624 **83**	A 21.500 **80**	C 26.000 **76**	F 25.000 **90**	D 23.125 **80**	B 25.250 **82**	G 21.250 **83**
7	B 23.250 **80**	E 19.624 **77**	G 20.500 **78**	C 23.250 **73**	A 19.500 **85**	F 20.750 **81**	D 22.500 **86**

Con estos resultados, realice lo siguiente:

a) Describa el modelo estadístico matemático, sin considerar la covariable.

b) Realice el ANOVA, emita sus conclusiones.

c) Calcule la Eficiencia Relativa (ER) del diseño cuadrado latino en comparación con el DCA , que es dada por la ecuación siguiente:

$$ER(\%) = \frac{CM_{Filas} + CM_{Columnas} + (t-1)\,CM_{Res\,DCL}}{(t+1)\,CM_{Res\,DCL}} \times 100$$

Recuerde que si el valor de ER(%) es menor o igual que 100%, se concluye que el diseño cuadrado latino no fue eficiente, en comparación con el DCA.

d) Calcule la Eficiencia Relativa (ER) del diseño cuadrado latino en comparación con el DBA , que es dada por la ecuación siguiente:

$$ER(\%) = \frac{CM_{Filas} + (t-2)\,CM_{Res\,DCL}}{(t-1)\,CM_{Res\,DCL}} \times 100$$

Recuerde que si el valor de ER(%) es > 100%, se concluye que el diseño cuadrado latino es más eficiente que el DBA.

e) Describa el modelo estadístico matemático, considerando la covariable.

f) Realice el ANOVA, considerando el modelo descrito en el inciso anterior. Emita sus conclusiones.

g) En caso de ser necesario, realice una prueba de comparación múltiple de medias.

h) Realice ambos análisis, usando SAS o Infostat.

7. Planifique un experimento donde incluya el uso de covariables. Justifique el diseño experimental utilizado, la covariable a medir, etc.

8. Investigue cuál es la importancia de la Covarianza en la experimentación agronómica, específicamente en el área de mejoramiento genético vegetal.

9. Defina Covarianza Múltiple, cite ejemplos de aplicación en la experimentación agropecuaria.

CAPÍTULO 11
ANÁLISIS DE GRUPOS DE EXPERIMENTOS

11.1 INTRODUCCIÓN

En la experimentación agronómica es común que ocurra la instalación de un grupo de experimentos, todos ellos con la misma estructura, sin embargo instalados en localidades distintas, con el objetivo de obtener conclusiones válidas para toda una región, admitiéndose siempre, que el efecto de las localidades es aleatorio.

Para cada experimento o para cada localidad se pueden obtener conclusiones analizando los datos individualmente y las conclusiones más generales serán obtenidas del análisis conjunto de todo el grupo de experimentos.

El agrupamiento de los experimentos para un análisis conjunto, según Campos (1984), podrá obedecer a diferentes criterios, dentro de los cuales se tiene:

a) por sectores geográficos, o
b) por año agrícolas, etc.

Para que los experimentos puedan ser reunidos es necesario que existe homogeneidad de varianzas, esto es, que los cuadrados medios del residuo no difieran mucho entre si. Para evaluar esa homogeneidad se puede utilizar cualquiera de las pruebas para verificación de la homogeneidad de varianzas, por ejemplo, la prueba de $F_{(máx)}$, también conocida como Prueba de Hartley. Si la homogeneidad de varianzas no fuera detectada, Nogueira (2000) indica que se debe proceder de la siguiente manera:

a) Separar en subgrupos los experimentos con cuadrados medios residuales semejantes, o no muy discrepantes y analizarlos separadamente, o

b) Realizar el análisis de varianza conjunto, y en el momento de tomar la decisión de rechazar o no la hipótesis nula, a través de la prueba de F, aplicar el método propuesto por Cochran en 1954, que consiste en ajustar los grados de libertad del residuo medio, representado por n, y los grados de libertad de la interacción Localidades × Tratamientos, representado por n´, obtenidos de la siguiente manera:

$$n = \frac{\left[CM_{Residuo(1)} + \ldots + CM_{Residuo(K)}\right]^2}{\dfrac{\left[CM_{Residuo(1)}\right]^2}{GL_{Residuo(1)}} + \ldots + \dfrac{\left[CM_{Residuo(K)}\right]^2}{GL_{Residuo(K)}}}, \text{ con } GL_{Residuo(k)} < n < \sum_{k=1}^{K} GL_{Residuo(k)},$$

$$n´ = \frac{(t-1)(K-1)^2 V_1^2}{(K-2)V_2 + V_1^2}, \qquad V_1 = \frac{\sum_{k=1}^{K} CM_{Residuo(k)}}{k}, \qquad V_2 = \frac{\sum_{k=1}^{K} \left[CM_{Residuo(k)}\right]^2}{k},$$

Siendo:
t es el número de tratamientos,
K es el número de localidades (o experimentos).
$CM_{Residuo(k)}$ es el cuadrado medio del residuo, referente a la k-ésima localidad o experimento.

11.2 EJEMPLO ILUSTRATIVO

Los datos que se presentan a continuación se refieren a la producción de caña de azúcar por hectárea (TCH), obtenidos de tres experimentos sobre evaluación de seis productos madurantes, instalados en tres fincas del ingenio "Palo Gordo", siguiendo un diseño en bloques al azar.

Finca Palo Gordo

Productos	Bloques				$Y_{i.1}$	Media
	I	II	III	IV		
Roundup	96.97	93.94	90.91	96.97	378.79	94.70
Roundup Max	96.97	96.97	90.91	96.97	381.82	95.45
Touchdown	102.00	93.94	93.94	93.94	383.82	95.95
Glifolaq	96.97	93.94	93.94	93.94	378.79	94.70
Fusilade	96.97	90.91	93.94	93.94	375.76	93.94
Select	96.97	90.00	93.94	96.97	377.88	94.47
Testigo	103.03	100.00	100.00	103.03	406.06	101.52
Totales	689.88	659.70	657.58	675.76	2,682.91	95.82

Finca Los Patos

Productos	Bloques				$Y_{i.2}$	Media
	I	II	III	IV		
Roundup	66.67	63.64	66.67	66.67	263.64	65.91
Roundup Max	63.64	63.64	66.67	60.61	254.55	63.64
Touchdown	63.64	63.64	66.67	66.67	260.61	65.15
Glifolaq	66.67	72.73	63.64	60.61	263.64	65.91
Fusilade	66.67	66.67	60.61	69.70	263.64	65.91
Select	63.64	60.61	63.64	66.67	254.55	63.64
Testigo	69.70	66.67	66.67	66.67	269.70	67.42
Totales	460.61	457.58	454.55	457.58	1,830.30	65.37

Finca La Reforma

Productos	Bloques				$Y_{i.3}$	Media
	I	II	III	IV		
Roundup	72.73	69.70	72.73	69.70	284.85	71.21
Roundup Max	72.73	72.73	81.82	66.67	293.94	73.48
Touchdown	69.70	66.67	72.73	72.73	281.82	70.45
Glifolaq	72.73	72.73	69.70	72.73	287.88	71.97
Fusilade	69.70	72.73	69.70	75.76	287.88	71.97
Select	69.70	72.73	75.76	72.73	290.91	72.73
Testigo	75.76	75.76	75.76	78.79	306.06	76.52
Totales	503.03	503.03	518.18	509.09	2,033.33	72.62

11.2.1 Análisis de varianza individual (por experimento o localidad)

El modelo adoptado fue el siguiente:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad \begin{cases} i = 1, 2, 3, \ldots, t \\ j = 1, 2, 3, \ldots, r \end{cases}$$

Siendo:

Y_{ij} = toneladas de caña por hectárea referentes al i-ésimo producto madurante en el el j-ésimo bloque o repetición.

μ = media general (TCH) del experimento.

τ_i = efecto del i-ésimo producto madurante

β_j = efecto del j-ésimo bloque o repetición (considerado como de efecto aleatorio)

ε_{ij} = error asociado a la ij-ésima unidad experimental.

Las hipótesis evaluadas son las siguientes:

Ho: $\mu_{1.} = \ . \ . \ . \ = \mu_{t.}$

Ha: por lo menos una $\mu_i \neq \mu_{i'}$, para $i \neq i'$,

usando la prueba de F y aplicando el análisis de varianza, cuyos resultados obtenidos para cada localidad o experimento, son los siguientes:

L_1 – Finca Palo Gordo

FV	GL	SC	CM	Valor de F	F crítica	P-value
Bloques	3	98.1021				
Madurantes	6	161.8751	26.9792	6.8763 **	2.66	0.00064025
Residuo	18	70.6234	3.9235			
Total	27	330.6006				

L_2 – Finca Los Patos

FV	GL	SC	CM	Valor de F	F crítica	P-value
Bloques	3	2.6236				
Madurantes	6	44.6019	7.4336	0.73913 ^{NS}	2.66	0.62512643
Residuo	18	181.0311	10.0573			
Total	27	228.2566				

L_3 – Finca La Reforma

FV	GL	SC	CM	Valor de F	F crítica	P-value
Bloques	3	21.973				
Madurantes	6	93.795	15.6325	1.50526 ^{NS}	2.66	0.23244321
Residuo	18	186.934	10.3852			
Total	27	302.702				

Como se observa en los cuadros anteriores, solamente en la finca Palo Gordo se detectaron diferencias significativas en el efecto de los productos utilizados como madurantes. Para poder reunir este grupo de experimentos y realizar el análisis de varianza conjunto es necesario verificar si ocurre la homocedasticidad entre los experimentos, a través de la aplicación de una prueba de homocedasticidad. Las hipótesis a evaluar son:

Ho : $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$

Ha : $\sigma_k^2 \neq \sigma_{k'}^2$, para $k \neq k'$, a través de la prueba de Hartley, cuyo resultado obtenido es,

$$F_{(máx)} = \frac{CM_{Residuo(máx)}}{CM_{Residuo(mín)}} = \frac{10.3852}{3.9235} = 2.65,$$

De la Tabla 5 del Apéndice se tiene para (0.05;3,18) = 3.19, por lo que no se rechaza la hipótesis nula, esto es, se puede considerar la existencia de homocedasticidad de varianzas, de tal manera, que es posible reunir en un análisis conjunto este grupo de experimentos.

11.2.2 Análisis conjunto

El modelo adoptado fue el siguiente:

$Y_{ijk} = \mu + \tau_i + \beta_{j/k} + \ell_k + (\tau\ell)_{ik} + \varepsilon_{ijk}$

$$\begin{cases} i = 1, 2, \ldots, t \text{ (madurantes)} \\ j = 1, 2, \ldots, r \text{ (bloques)} \\ k = 1, 2, \ldots, K \text{ (localidades o experimentos)} \end{cases}$$

Siendo:

Y_{ijk} = toneladas de caña por hectárea referentes al i-ésimo producto madurante en el el j-ésimo bloque o repetición de la k-ésima localidad;

μ = media general

τ_i = efecto del i-ésimo producto madurante

$\beta_{j/k}$ = efecto del j-ésimo bloque en la k-ésima localidad,

ℓ_k = efecto de la k-ésima localidad,

$(\tau\ell)_{ik}$ = efecto de la interacción entre el i-ésimo producto madurante y la k-ésima localidad,

ε_{ijk} = error experimental asociado a la observación Y_{ijk}.

Barbin (2013) cita que, el efecto de bloques dentro de localidades ($\beta_{j/k}$) no representa importancia práctica, por lo que puede ser eliminado del modelo. Sin embargo, si en algunos tipos de experimentos ese efecto fuera importante, se puede admitir el modelo completo. Las hipótesis a ser evaluadas son:

i) H_{o1}: $\tau_i = 0$, para todo i, contra,
H_{a1}: $\tau_i \neq 0$, para algún i;

ii) H_{o1}: $\ell_k = 0$, para todo k, contra,
H_{a1}: $\ell_k \neq 0$, para algún k;

iii) H_{o1}: $(\tau\ell)_{ik} = 0$, para todo i y k, contra,
H_{a1}: $(\tau\ell)_{ik} \neq 0$, para algún i y k.

De esta manera se tienen las siguientes fuentes de variación y el cálculo de los grados de libertad:

Fuentes de variación	**Grados de libertad**
Bloques dentro de localidades	K (r – 1)
Localidades	(K – 1)
Tratamientos	(t – 1)
Localidades × Tratamientos	(K – 1) (t – 1)
Residuo (error experimental)	K (t – 1) (r – 1)
Total	(t r K – 1)

Para realizar el análisis de varianza conjunto, primero se calcula el gran total así:

$$Y... = \sum_{i=1_}^{t}\sum_{j=1}^{r}\sum_{k=1}^{K} Y_{ijk} = 96.97 + 93.94 + ... + 78.79 = 6546.73$$

El factor de corrección (FC) = $FC = \dfrac{Y...^2}{(trK)} = \dfrac{(6546.73)^2}{(7 \times 4 \times 3)} = 510234.211$

La suma de cuadrados del total (SC_{tot}) se obtiene de la siguiente manera:

$$SC_{tot} = \sum_{i=1_}^{t}\sum_{j=1}^{r}\sum_{k=1}^{K} Y_{ijk}{}^2 - FC = (96.97^2 + 93.94^2 + ... + 78.79^2) - 510234.211 = 15026.4805$$

Para el cálculo de las otras sumas de cuadrados, debemos organizar un cuadro auxiliar que relaciona los productos aplicados y los experimentos, o sea:

	Localidades			
Productos	1	2	3	Yi..
Roundup	378.79	263.65	284.86	927.3
Roundup Max	381.82	254.56	293.95	930.33
Touchdown	383.82	260.62	281.83	926.27
Glifolaq	378.79	263.65	287.89	930.33
Fusilade	375.76	263.65	287.89	927.3
Select	377.88	254.56	290.92	923.36
Testigo	406.06	269.71	306.07	981.84
Y..k	2682.92	1830.4	2033.41	6546.73

De ese cuadro obtenemos:

$$SC_{trats} = \frac{\sum_{i=1}^{t} Y_{i..}{}^2}{rK} - FC = \frac{927.3^2 + ... + 981.84^2}{(4 \times 3)} - FC = 213.954862$$

$$SC_{localidades} = \frac{\sum_{i=1}^{t} Y_{..k}{}^{2}}{tr} - FC = \frac{2682.92^2 + 1830.4^2 + 2033.41^2}{(7 \times 4)} - FC = 14165.0792$$

$$SC_{trats,localidades} = \frac{\sum_{i=1}^{t} Y_{i.k}{}^{2}}{r} - FC = \frac{378.79^2 + 263.65^2 + ... + 306.07^2}{4} - FC = 14465.2939$$

$$SC_{trats \times localidades} = SC_{trats,localidades} - SC_{trats} - SC_{localidades}$$
$$= 14465.2939 - 213.954862 - 14165.0792 = 86.2599095$$

$$SC_{bloques(localidades)} = SC_{bloques(localidad1)} + SC_{bloques(localidad2)} + SC_{bloques(localidad3)}$$

$$SC_{bloques(localidades)} = 98.1021 + 2.6236 + 21.973 = 122.6987$$

La suma de cuadrados del residuo del análisis conjunto es la suma de los cuadrados de los residuos de los análisis individuales, esto es:

$$SC_{residuo} = SC_{residuo1} + SC_{residuo2} + SC_{residuo3}$$

$$SC_{residuo} = 70.6234 + 181.0311 + 186.934 = 438.5885$$

El análisis de varianza conjunto de los experimentos se presenta a continuación:

FV	GL	SC	CM	Valor de F	F crítica
Bloques (Localidades)	9	122.6987	13.633189		
Localidades	2	14165.0792	7082.5396	519.507186**	4.26
Tratamientos	6	213.95486	35.659143	4.39043372*	2.27
Localidades ×Tratamientos	12	86.2599095	7.1883258	0.8850428	1.94
Residuo	54	438.5885	8.122009		
Total	83	330.6006			

Conclusiones:

El mayor efecto significativo observado se refiere a las localidades. Los madurantes también presentan diferencias significativas, pero no así la interacción. Por tanto los factores pueden analizarse por separado.

*****

Nota:
Para resolver este ejemplo en Infostat, especifique en el modelo los siguientes efectos:

Local \ Local>Rep
Local>Rep
Mad
Local*Mad

11.3 PROGRAMA EN SAS PARA UNA SERIE DE EXPERIMENTOS

```
OPTIONS nodate nonumber;
TITLE "Serie de experimentos en bloques al azar";
DATA cana;
INPUT local$ mad$ rep tch;
CARDS;
PG      Rup     1       96.97
PG      Rmax    1       96.97
PG      Tdown   1       102
…  …  …  …
…  …  …  …
…  …  …  …
LR      Glif    4       72.73
LR      Fusil   4       75.76
LR      Sel     4       72.73
LR      Test    4       78.79
;
PROC PRINT; title "Datos para verificación"; RUN;
/*Análisis para cada localidad*/
PROC SORT;  by local;
RUN;
PROC GLM; title "Análisis individuales";  by local;
CLASS mad rep;
MODEL tch = rep mad/SS1;
MEANS mad/tukey;
RUN;
/*Análisis de Varianza Conjunta*/
PROC GLM; title 'Análise conjunta I';
CLASS local mad rep;
MODEL tch = local rep(local) mad mad*local/SS1;
TEST H=local E=rep(local);
MEANS mad/tukey;
MEANS local/tukey E=rep(local);
RUN;
PROC GLM; title 'Análise conjunta II';
CLASS local mad rep;
MODEL tch = local rep(local) mad mad*local/SS1;
LSMEANS mad*local/slice=local;
RUN;
```

Un resumen de la salida que genera este programa se presenta a continuación:

```
                                    Análisis individuales

---------------------------------------------- local=LP --------------------------------------

                                      The GLM Procedure
                                   Class Level Information

                     Class         Levels    Values

                     mad                7    Fusil Glif Rmax Rup Sel Tdown Test
                     rep                4    1 2 3 4
```

Number of Observations Read 28
Number of Observations Used 28

The GLM Procedure

Dependent Variable: tch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	47.2160571	5.2462286	0.52	0.8401
Error	18	180.9948857	10.0552714		
Corrected Total	27	228.2109429			

R-Square	Coeff Var	Root MSE	tch Mean
0.206897	4.850750	3.171005	65.37143

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	3	2.62311429	0.87437143	0.09	0.9663
mad	6	44.59294286	7.43215714	0.74	0.6251

Análisis individuales

-- **local=LR** ---------------------------------------

The GLM Procedure

Class Level Information

Class	Levels	Values
mad	7	Fusil Glif Rmax Rup Sel Tdown Test
rep	4	1 2 3 4

Number of Observations Read 28
Number of Observations Used 28

The GLM Procedure

Dependent Variable: tch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	115.7449179	12.8605464	1.24	0.3328
Error	18	186.8968929	10.3831607		
Corrected Total	27	302.6418107			

R-Square	Coeff Var	Root MSE	tch Mean
0.382449	4.437086	3.222291	72.62179

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	3	21.96858214	7.32286071	0.71	0.5612
mad	6	93.77633571	15.62938929	1.51	0.2324

Análisis individuales

-- local=PG ---------------------------------------

The GLM Procedure

Class Level Information

Class	Levels	Values
mad	7	Fusil Glif Rmax Rup Sel Tdown Test
rep	4	1 2 3 4

Number of Observations Read 28
Number of Observations Used 28

The GLM Procedure

Dependent Variable: tch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	259.9321786	28.8813532	7.36	0.0002
Error	18	70.6163643	3.9231313		
Corrected Total	27	330.5485429			

R-Square	Coeff Var	Root MSE	tch Mean
0.786366	2.067125	1.980690	95.81857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	3	98.0866857	32.6955619	8.33	0.0011
mad	6	161.8454929	26.9742488	6.88	0.0006

The GLM Procedure

Tukey's Studentized Range (HSD) Test for tch

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	3.923131
Critical Value of Studentized Range	4.67313
Minimum Significant Difference	4.628

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	mad
A	101.515	4	Test
B	95.955	4	Tdown
B	95.455	4	Rmax
B	94.698	4	Glif
B	94.698	4	Rup
B	94.470	4	Sel
B	93.940	4	Fusil

Análise conjunta I

The GLM Procedure

Class Level Information

Class	Levels	Values
local	3	LP LR PG
mad	7	Fusil Glif Rmax Rup Sel Tdown Test
rep	4	1 2 3 4

Number of Observations Read	84
Number of Observations Used	84

Análise conjunta I

The GLM Procedure

Dependent Variable: tch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	14587.97233	503.03353	61.95	<.0001
Error	54	438.50814	8.12052		
Corrected Total	83	15026.48047			

R-Square	Coeff Var	Root MSE	tch Mean
0.970818	3.656342	2.849653	77.93726

Source	DF	Type I SS	Mean Square	F Value	Pr > F
local	2	14165.07917	7082.53959	872.18	<.0001
rep(local)	9	122.67838	13.63093	1.68	0.1170
mad	6	213.95486	35.65914	4.39	0.0011
local*mad	12	86.25991	7.18833	0.89	0.5666

Tests of Hypotheses Using the Type I MS for rep(local) as an Error Term

Source	DF	Type I SS	Mean Square	F Value	Pr > F
local	2	14165.07917	7082.53959	519.59	<.0001

Tukey's Studentized Range (HSD) Test for tch

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	54
Error Mean Square	8.120521
Critical Value of Studentized Range	4.33055
Minimum Significant Difference	3.5624

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	mad
A	81.820	12	Test
B	77.528	12	Glif
B	77.528	12	Rmax
B	77.275	12	Rup
B	77.275	12	Fusil
B	77.189	12	Tdown
B	76.947	12	Sel

Tukey's Studentized Range (HSD) Test for tch

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	9
Error Mean Square	13.63093
Critical Value of Studentized Range	3.94850
Minimum Significant Difference	2.755

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	local
A	95.8186	28	PG
B	72.6218	28	LR
C	65.3714	28	LP

Análise conjunta II

The GLM Procedure

Class Level Information

Class	Levels	Values
local	3	LP LR PG
mad	7	Fusil Glif Rmax Rup Sel Tdown Test
rep	4	1 2 3 4

Number of Observations Read	84
Number of Observations Used	84

Análise conjunta II
The GLM Procedure

Dependent Variable: tch

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	14587.97233	503.03353	61.95	<.0001
Error	54	438.50814	8.12052		
Corrected Total	83	15026.48047			

R-Square	Coeff Var	Root MSE	tch Mean
0.970818	3.656342	2.849653	77.93726

Source	DF	Type I SS	Mean Square	F Value	Pr > F
local	2	14165.07917	7082.53959	872.18	<.0001
rep(local)	9	122.67838	13.63093	1.68	0.1170
mad	6	213.95486	35.65914	4.39	0.0011
local*mad	12	86.25991	7.18833	0.89	0.5666

Análise conjunta II
The GLM Procedure
Least Squares Means

local	mad	tch LSMEAN
LP	Fusil	65.912500
LP	Glif	65.912500
LP	Rmax	63.640000
LP	Rup	65.912500
LP	Sel	63.640000
LP	Tdown	65.155000
LP	Test	67.427500
LR	Fusil	71.972500
LR	Glif	71.972500
LR	Rmax	73.487500
LR	Rup	71.215000
LR	Sel	72.730000
LR	Tdown	70.457500
LR	Test	76.517500
PG	Fusil	93.940000
PG	Glif	94.697500
PG	Rmax	95.455000
PG	Rup	94.697500
PG	Sel	94.470000
PG	Tdown	95.955000
PG	Test	101.515000

local*mad Effect Sliced by local for tch

local	DF	Sum of Squares	Mean Square	F Value	Pr > F
LP	6	44.592943	7.432157	0.92	0.4912
LR	6	93.776336	15.629389	1.92	0.0934
PG	6	161.845493	26.974249	3.32	0.0074*

****

Para mayor infomación sobre uso del Infostat en experimentación agronómica, puede consultar el documento siguiente:

Morales, J.; Quemé, JL; Melgar, M. 2009. **Infostat. Manual de uso: Ejemplos de los principales métodos estadísticos utilizados en la industria cañera.** Santa Lucía Cotz.: CENGICAÑA. 48 p. Disponible en: http://www.cengicana.org/descargas/ManualInfoStat.pdf

11.4 EJERCICIOS PROPUESTOS

1. Los datos que se presentan a continuación se refieren a las alturas (en metros, promedio de 25 plantas por parcela) de árboles de *Eucaliptus grandis*, con 7 años de edad (en 1980) de tres ensayos en bloques completos al azar, sobre 6 tratamientos (progenies), citado por Barbin (2013)

Ensayo en Araraquara

Tratamientos	Bloques				Total
	I	II	III	IV	
T_1	20.3	19.6	23.5	19.1	82.5
T_2	21.7	19.3	16.7	18.5	76.2
T_3	22.0	24.9	24.4	20.8	92.1
T_4	20.8	23.0	21.3	24.9	90.0
T_5	21.5	22.3	22.1	21.9	87.8
T_6	19.6	17.7	18.7	22.0	78.0
Totales	125.9	126.8	126.7	127.2	506.6

Ensayo en Bento Quirino

Tratamientos	Bloques				Total
	I	II	III	IV	
T_1	10.2	11.7	9.1	8.1	39.1
T_2	16.1	10.8	10.9	10.3	48.1
T_3	17.7	13.1	14.2	11.0	56.0
T_4	13.5	14.4	11.2	12.8	51.9
T_5	20.5	12.5	11.3	12.2	56.5
T_6	12.0	13.0	12.3	10.6	47.9
Totales	90.0	75.5	69.0	65.0	299.5

Ensayo en Mogi-Guaçú

Tratamientos	Bloques				Total
	I	II	III	IV	
T_1	22.7	21.4	22.9	22.0	89.0
T_2	22.6	21.4	20.7	20.8	85.5
T_3	21.4	21.7	22.5	19.4	85.0
T_4	25.0	23.6	23.3	24.8	96.7
T_5	26.4	26.4	28.0	27.3	108.1
T_6	20.6	23.5	19.4	21.9	85.4
Totales	138.7	138.0	136.8	136.2	549.7

Fuente: Instituto Florestal –Tupi, SP

T_1 :	Pretoria (Procedente de Sudáfrica)	T_4 :	2094 (Progenie Rio Claro)
T_2 :	637 (Progenie Rio Claro)	T_5 :	9559 (Procedente de Australia)
T_3 :	2093 (Progenie Rio Claro)	T_6 :	9575 (Procedente de Australia)

a) Realice los análisis de varianza individuales (para cada localidad) y concluya.
b) Verifique la relación entre el mayor y menos CMee y confirme si los ensayos pueden ser reunidos en un único análisis conjunto, sin restricciones.
c) Realice el análisis conjunto de los experimentos.
d) Efectúe el desdoblamiento de los grados de libertad relativos a los tratamientos + Interacción Tratamiento x Localidad.

2. Los datos de productividad de naranja, en kg/planta (promedio de 2 plantas), se refieren a un ensayo en bloques completos al azar de distanciamientos de naranja Valencia sobre *trifoliata* realizado en la Estación Experimental de Limeira del Instituto Agronómico de Campinas (IAC). Los distanciamientos (en m) utilizados fueron: 6 x 2, 6 x 3, 6 x 4, 6 x 5 y 6 x 6. Las cosechas fueron realizadas de 1975 a 1979.

Años	Distanciamiento	Bloques			
		I	II	III	IV
1975	6 x 2	9.0	5.25	8.70	3.50
	6 x 3	4.25	2.25	2.60	7.35
	6 x 4	7.00	5.10	8.00	3.85
	6 x 5	5.50	4.80	3.30	6.45
	6 x 6	4.05	3.85	9.50	3.75
1976	6 x 2	18.25	21.80	20.40	20.25
	6 x 3	16.30	11.15	12.35	24.10
	6 x 4	21.40	17.55	19.15	18.35
	6 x 5	17.05	13.35	15.10	18.90
	6 x 6	19.30	14.15	24.30	13.25
1977	6 x 2	32.50	16.65	22.70	8.75
	6 x 3	23.05	26.10	21.95	25.60
	6 x 4	18.05	23.30	27.55	26.80
	6 x 5	22.05	21.25	16.25	15.30
	6 x 6	28.15	28.55	29.10	18.50
1978	6 x 2	48.90	56.75	48.20	68.40
	6 x 3	55.35	41.85	47.15	54.35
	6 x 4	60.60	57.70	61.20	53.05
	6 x 5	51.80	59.45	48.05	58.45
	6 x 6	70.50	58.55	67.00	49.35
1979	6 x 2	12.00	23.15	26.35	19.15
	6 x 3	35.35	24.25	36.75	19.75
	6 x 4	17.75	25.00	35.00	21.75
	6 x 5	34.00	11.90	32.00	8.50
	6 x 6	27.50	29.85	22.50	22.75

a) Realice el análisis conjunto de los experimentos.
b) Efectúe el desdoblamiento de los grados de libertad relativos a los años + Interacción años x distanciamiento.
c) Haga el análisis de regresión hasta 2º grado para años dentro de cada distanciamiento. Establezca las ecuaciones de regresión e obtenga los coeficientes de determinación.

3. Gómez Pereira, P. (2006) realizó la investigación titulada: “Evaluación de dos inhibidores de floración de caña de azúcar (*Saccharum* spp) en tres localidades de la zona cañera del ingenio Pantaleón”. El diseño experimental utilizado fue bloques al azar, con tres repeticiones. Una de las variables de respuesta medidas fue el rendimiento expresado en kilogramos de azúcar por tonelada de caña cortada. La variedad de caña utilizada fue CP-722086. Los resultados se presentan en la siguiente tabla:

		Bloques		
Localidades	Producto	1	2	3
	Ethrel	130.59	128.04	135.54
Lote Petapilla, Finca Limones	Optilux	136.62	137.09	132.21
	Testigo	103.74	141.8	130.69
	Ethrel	138.46	132.1	135.56
Lote San Joaquin, Finca San Bonifacio	Optilux	144.52	139.86	134.16
	Testigo	133.68	133.04	133.04
	Ethrel	139.31	138.73	138.73
Lote El Común, Finca San Bonifacio	Optilux	141.28	130.59	103.74
	Testigo	136.95	131.11	135.63

a) Realice los análisis de varianza individuales (para cada localidad) y concluya.
b) Verifique la relación entre el mayor y menos CMee y confirme si los ensayos pueden ser reunidos en un único análisis conjunto, sin restricciones.
c) Realice el análisis conjunto de los experimentos.
d) Efectúe el desdoblamiento de los grados de libertad relativos a los tratamientos + Interacción Tratamiento × Localidad.

4. Un experimento donde se evaluaron tres variedades de arroz (4440, CICA-9 y Blue Bonnet-50) en tres localidades de Olancho (Concepción, La Bomba, Las Delicias y el Campanario), Honduras fue realizado utilizando un diseño de bloques al azar en cada localidad. Los resultados se expresaron en toneladas métricas por hectárea.

Lugar	Bloque	Variedad	Rend
Concepción	1	4440	4.40
Concepción	2	4440	5.60
Concepción	3	4440	5.20
Concepción	4	4440	4.10
Concepción	1	CICA-9	3.83
Concepción	2	CICA-9	5.00
Concepción	3	CICA-9	4.30
Concepción	4	CICA-9	4.50
Concepción	1	BB-50	4.60
Concepción	2	BB-50	4.90
Concepción	3	BB-50	3.80

Concepción	4	BB-50	4.10
Bomba	1	4440	4.61
Bomba	2	4440	5.20
Bomba	3	4440	5.10
Bomba	4	4440	5.30
Bomba	1	CICA-9	4.20
Bomba	2	CICA-9	5.25
Bomba	3	CICA-9	4.40
Bomba	4	CICA-9	4.75
Bomba	1	BB-50	3.59
Bomba	2	BB-50	4.50
Bomba	3	BB-50	5.10
Bomba	4	BB-50	5.60
Delicias	1	4440	9.10
Delicias	2	4440	6.99
Delicias	3	4440	8.45
Delicias	4	4440	7.15
Delicias	1	CICA-9	7.41
Delicias	2	CICA-9	7.01
Delicias	3	CICA-9	6.50
Delicias	4	CICA-9	6.18
Delicias	1	BB-50	4.00
Delicias	2	BB-50	4.92
Delicias	3	BB-50	4.50
Delicias	4	BB-50	3.72
Campanario	1	4440	4.50
Campanario	2	4440	5.50
Campanario	3	4440	4.75
Campanario	4	4440	6.00
Campanario	1	CICA-9	6.00
Campanario	2	CICA-9	5.35
Campanario	3	CICA-9	4.70
Campanario	4	CICA-9	5.10
Campanario	1	BB-50	4.70
Campanario	2	BB-50	4.00
Campanario	3	BB-50	3.70
Campanario	4	BB-50	4.20

BIBLIOGRAFÍA

1. Alcarde, R. Fundamentos do diagrama de Hasse e aplicações à experimentação. 2007. Dissertação (Mestrado), Brasil. Universidad de Sao Paulo, Escola Superior de Agricultura "Luiz de Queiroz". 99 p.

2. Alfaro Argueta, RA. 1999. Evaluación inicial del efecto de tres intensidades de raleo y tres de poda en el crecimiento de una plantación de *Pinus caribaea* Morelet var. *hondurensis*, en Livingston, Izabal Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 124 p.

3. Alfaro Villatoro, MA. 1985. Evaluación del rendimiento y composición química del amaranto *(Amaranthus hypochondriacus* L.) en tres diferentes épocas de corte. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 48 p.

4. Álvarez Cajas, VM. 1982. Determinación del tamaño óptimo de parcela experimental en caña de azúcar *(Saccharum officinarum* L.) bajo condiciones de la finca Bulbuxyá. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 49 p.

5. Anjos, A. Dos 2003. Planejamento de experimentos I. Curitiba: Universidade Federal do Paraná, Setor de Ciências Exatas, Departamento de Estatística. 94 p.

6. Ayala Marroquín, ME. 1992. Efecto de cuatro frecuencias de riego en el rendimiento y la evapotranspiración del maíz *(Zea mays* L.), en la Unidad de Riego San Cristóbal Acasaguastlán. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 101 p.

7. Banzatto, DA.; Kronka, S. 2011. Experimentação Agrícola. 3ª Impressão, 4ª Ed. Jabuticabal: Funep. 237 p.

8. Barbin, D. 2013. Planejamento e análise estatística de experimentos agronômicos. Londrina: Mecenas. 214 p.

9. Barrientos García, M. 1981. Evaluación de cuatro métodos para la determinación de tamaño y forma óptimos de parcela para experimentos agrícolas. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 60 p.

10. Batres, J.; De León, WE; Salazar, R.; Saquimux, F. 2000. Efecto de la edad de los árboles semilleros en el porcentaje de germinación de semilla de pinabete *(Abies guatemalensis* Rehder) en la comunidad de Chimente, Totonicapán, Guatemala. Ministerio de Agricultura, Ganadería y Alimentación, Área de Recursos Naturales Renovables, Subárea de Forestería, Región VII. 35 p.

11. Benítez, C.; Pece, M.; Galindez, M. 2010. Análisis de la Varianza en experimentos factoriales. Universidad Nacional de Santiago del Estero (Argentina), Facultad de Ciencias Forestales. Série Didática No. 21. 47 p.

12. Campbell, D. y Stanley, J. 1982. Diseños experimentales y cuasiexperimentales en investigación social. Buenos Aires: Amorrortu Editores.

13. Campos, H. de. 1984. Estatística aplicada à experimentação com cana-de-açúcar. Piracicaba: FEALQ. 292 p.

14. Castillo Fratti, AJ. 1996. Evaluación del efecto de la incisión anular sobre la calidad y rendimiento de la fruta de uva de mesa *(Vitis vinifera* L.) en dos localidades del nororiente del país. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 61 p.

15. Castaño, E.; Domínguez, J. 2001. Diseño de experimentos para el Desarrollo Tecnológico y Mejora Industrial. México, D.F.: Just in Time Press, S.A. de C.V. 313 p.

16. CENGICAÑA. 2005. Memoria. Presentación de resultados de investigación. Zafra 2004–2005. Guatemala. 185 p.

17. Chacin, FL. 1976. Tamaño de parcela experimental y su forma. Revista Facultad de Agronomía (Maracay), 60(3):55-74.

18. Di Rienzo, JA. *et al*. 2009. Estadística para las Ciencias Agropecuarias. 7ª Ed. Córdoba: Brujas. 356 p.

19. Di Rienzo, JA. *et al*. InfoStat versión **2015**. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL http://www.infostat.com.ar

20. Estrada Muy, RA. 1993. Evaluación del efecto de 16 distancias de siembra sobre el crecimiento y rendimiento en el cultivo de chile chocolate *(Capsicum* sp) en el valle central de Guatemala. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 48 p.

21. Garcia, AAF.; Barbin, D.; Piedade, S. Apostila Aulas Estatística Experimental com SAS. Publicação didática. Departamento de Ciências Exatas, ESALQ/USP, Piracicaba, SP. Março de 2002. 65 p.

22. Gomes, FP. 2009. Curso de Estatística Experimental. 15 Ed. Piracicaba: FEALQ. 451 p.

23. González Ramírez, BH. 1999. Evaluación agroeconómica del sistema de cultivo intercalado caña de azúcar (Saccharum spp.)–girasol *(Helianthus agnus* L.) Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 59 p.

24. Hernández Dávila, AG. 1982. Determinación del tamaño óptimo de parcela para estudios experimentales en dos variedades de papa *(Solanum tuberosum* L.) en el altiplano central de Guatemala. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 58 p.

25. Herrera, J.; Barreras, A. 2001. Manual de procedimientos: Análisis estadístico de experimentos pecuarios (utilizando el programa SAS). 2ª. Ed. Montecillo: Colegio de Postgraduados. 117 p.

26. Instituto Interamericano de Cooperación para la Agricultura (IICA). 1988. Notas del curso taller: Metodología de la Investigación Pecuaria: Diseños Experimentales. Comayagua, Honduras. 3 al 14 de agostos de 1987. 141 p.

27. Jayaraman, K. 2000. A Statistical Manual for Forestry Research. Bangkok: FORSPA-FAO Publication. 240 p.

28. Kempthorne, O. 1975. The Design and Analysis of Experiments. New York: Krieger. 631 p.

29. López Bautista, EA. 1999. Influencia de la lámina de riego en el efecto del madurante en caña de azúcar *(Saccharum* spp.) var. CP-722086 Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 70 p.

30. López Galindo, HR. 1984. Respuesta del arroz (*Oriza sativa* L.) a la aplicación 4 niveles de calcio y 3 dosis de fósforo en suelos ácidos serie Cristina. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 100 p.

31. López Valenzuela, MA. 1999. Evaluación de dos densidades de siembra y cuatro concentraciones de etephon sobre la brotación, población y producción de cinco variedades de caña de azúcar *(Saccharum officinarum* L.) en la Estación Experimental Camantulul, Santa Lucía Cotzumalguapa, Escuintla. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 66 p.

32. Macchiavelli, R. 2015. AGRO 6600 Curso avanzado de Biometría (en línea). Mayagüez, Puerto Rico. Consultado 27 En. 2015. Disponible en: http://academic.uprm.edu/rmacchia/agro6600/agro6600.pdf

33. Martínez Garza, A. 1994. Experimentación Agrícola. México: Universidad Autónoma de Chapingo. 342 p.

34. Martínez Garza, A. 1988. Diseños Experimentales: Métodos y Elementos de Teoría. México, D.F.: Trillas. 756 p.

35. Melo, O.; López, L.; Melo, S. 2007. Diseños de Experimentos: Métodos y Aplicaciones. Bogotá, D.C.: Universidad Nacional de Colombia. 544 p.

36. Mendoza Leonardo, AO. 2001. Evaluación del efecto de la pulpa del café *(Coffea arabica* L.) sobre el rendimiento y eficiencia biológica de la cepa ECS-0110 de *Pleurotas ostreatus* utilizando estopa de coco *(Cocus nucifera* L.) y estróbilos de pino *(Pinus* spp.) como sustratos. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 46 p.

37. Montgomery, D. 2004. Diseño y análisis de experimentos. 2a. Ed. México D.F.: Limusa Wiley. 692 p.

38. Navarro Estrada, JG. 2003. Efecto de cuatro láminas de riego sobre el rendimiento de plátano *(Musa paradisiaca* var. Curraré) bajo las condiciones de aldea Los Encuentros, Coatepeque, Quetzaltenango. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 64 p.

39. Navarro Sosa, MR. 1996. Evaluación de aceites y detergentes para el control de áfidos *(Homóptera: Aphididae)* en el cultivo de brócoli en Chimaltenango. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 59 p.

40. Neter, J.; Kutner, M.; Nachtsheim, C.; Wasserman, W. 1996. Applied statistical models. 4a. Ed. New York: McGraw-Hill. 1048 pp

41. Nogueira, MCS. 2007. Experimentação Agronómica I – Conceitos, planejamento e análise estatística. Piracicaba: MCS Nogueira. 479 p.

42. Orellana Najarro, SE. 2006. Evaluación de la selectividad de los herbicidas Acetoclor y Alaclor en seis cultivos hortícolas en el municipio de Monjas, Jalapa. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 86 p.

43. Ostle, B. 1973. Estadística aplicada. México D.F.: Limusa-Wiley. 629 p.

44. Peña Hernández, A. 1997. Evaluación de 14 líneas de frijol Tepary (*Phaseolus acutifolius* Gray) en tres localidades de El Progreso. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 123 p.

45. Quiroga, V. 1976. Manual práctico para el análisis de experimentos. San José, Costa Rica: IICA, Programa de Información Agropecuaria del Istmo Centroamericano. 113 p.

46. Quaggio, JA.; Mascarenhas, HA. Bataglia, OC. Resposta da soja à aplicação de doses crescentes de calcário en Latossolo roxo distrófico de cerrado. II. Efeito residual. Revista Brasileira de Ciência do Solo. 6:113-8, 1982.

47. Ramalho, MAP; Ferreira, DF, Oliveira, AC de. 2012. Experimentação em genética e melhoramento de plantas. 3ª Ed. Lavras: Editora UFLA. 328 p.

48. Resende, MDV. de 2007. Matemática e estatística na análise de experimentos no melhoramento genético. Colombo: EMBRAPA Florestas. 362 p.

49. Restrepo, L. Diagramas de estructura en el análisis de varianza. Revista Colombiana de Ciencias Pecuarias. 20: 200–208, 2007.

50. Rivera Aguirre, FA. 2001. Evaluación de N y P. en el rendimiento de cardamomo *(Elettaria cardamomun* M.), en la serie de suelos Tamahu, aldea Choval, Cobán, Alta Verapaz. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 56 p.

51. Salazar Rodríguez, JA. 1985. Determinación de la cantidad mínima efectiva y residualidad del trimedlure, para mejorar la eficiencia de detección de la mosca del mediterráneo *(Ceratitis capitata* Wiedemann) Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 53 p.

52. Santos, J.W. Dos *.et al*. 2008. Estatística Experimental Aplicada. 2ª Ed. Campina Grande: EMBRAPA Algodão/ Universidade Federal de Campina Grande. 461 p.

53. Sandoval García, CA . 1987. Evaluación de 2 densidades de siembra y 4 niveles de fertilización nitrogenada en 3 cultivares de bledo (*Amaranthus spp*.) en Villa Canales, Guatemala. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 44 p.

54. Sinha, SP. 2008. Diseño y Análisis de Experimentos usando SAS (en línea) Instituto de Estadística Aplicada y Computación, Universidad de Los Andes, Mérida, Venezuela. Consultado 25 En. 2015. Disponible en: http://webdelprofesor.ula.ve/economia/sinha/index.htm#beg

55. Sosa Leonardo, JE. 1999. Evaluación de cuatro substancias diluyentes-dispersantes de polen para producir semilla híbrida en cuatro cultivares de Marigold *(Tagetes erecta* L.) mediante polinización artificial en condiciones de invernadero. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 60 p.

56. Steel, RG; Torrie, JH. 1992. Bioestadística: principios y procedimientos. México: McGraw-Hill. 622 p.

57. Storck, L.*et al*. 2011. Experimentação vegetal. 3ª Ed. Santa María, RS: UFSM. 200 p.

58. Stroup, W. *et al*. 2012. Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences. Madison, WI: American Society of Agronomy, Soil Science Society of America and Crop Science Society of America. 283 p.

59. Vargas, VC. 2012. Evaluación de tres niveles, dos fuentes y dos fomas de aplicación del micronutriente hierro (Fe) en el cultivo de mango (*Mangifera indica* cv Tommy Atkins) en la finca "El Tintero", El Jícaro, departamento de El Progreso. Investigacion EPS. FAUSAC. 48 p.

60. Teos Morales, EA. 1980. Determinación del nivel de tolerancia de la planta de maíz (*Zea mays* L.) al daño causado por el gusano cogollero (larva de Spodoptera frugiperda J. E. Smith) en el parcelamiento La Máquina. Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 55 p.

61. Veras Castillo, EE. 1992. Evaluación de dos atrayentes sexuales y tres mezclas de éstos en captura de mosca de la fruta del mediterráneo *(Ceratitis capitata* Wied) Tesis Ing. Agr. Guatemala, Universidad de San Carlos, Facultad de Agronomía. 72 p.

APÉNDICE
TABLAS ESTADÍSTICAS

Tabla 1 Puntos porcentuales de la distribución F, $\alpha = 0.05$

Grados de libertad del denominador	Grados de libertad del numerador																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	161.4	199.5	215.7	224.6	230.2	233.9	236.8	238.9	240.54	241.9	242.9	243.9	244.7	245.4	245.9	246.5	246.9	247.3	247.7	248.0	248.3	248.6	248.8	249.0	249.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45	19.45	19.45	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.65	8.65	8.64	8.64	8.63
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.79	5.79	5.78	5.77	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56	4.55	4.54	4.53	4.53	4.52
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87	3.86	3.86	3.85	3.84	3.83
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.43	3.43	3.42	3.41	3.40
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.93	2.92	2.91	2.90	2.89
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77	2.76	2.75	2.75	2.74	2.73
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65	2.64	2.63	2.62	2.61	2.60
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.53	2.52	2.51	2.51	2.50
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.45	2.44	2.43	2.42	2.41
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39	2.38	2.37	2.36	2.35	2.34
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33	2.32	2.31	2.30	2.29	2.28
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.26	2.25	2.24	2.24	2.23
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.22	2.21	2.20	2.19	2.18
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19	2.18	2.17	2.16	2.15	2.14
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16	2.14	2.13	2.12	2.11	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.11	2.10	2.09	2.08	2.07
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.08	2.07	2.06	2.05	2.05
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06	2.05	2.04	2.02	2.01	2.01	2.00
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03	2.01	2.00	1.99	1.98	1.97
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02	2.01	2.00	1.98	1.97	1.96	1.96

Tabla 2 Valores de la amplitud total estudentizada, q 0.05, para uso en la prueba de Tukey

	Número de tratamientos																		
Glee	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83	59.56
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75	18.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.15	10.35	10.53	10.69	10.84	10.98	11.11	11.24
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.34	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
30	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.72	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
40	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.64	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
50	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.59	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.94	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

glee = grados de libertad del error experimental.

Tabla 3 Valores de la amplitud total estudentizada, d 0.05, para uso en la prueba de Duncan

	Número de tratamientos										
glee	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**20**	**50**
1	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.48	3.47
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.47	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.47	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.47	3.67

Tabla 4 Valores críticos para la prueba de Dunnett de comparación de tratamientos con un control, d0.05, comparaciones bilaterales

	Número de medias de tratamientos (excluyendo el control)								
glee	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.54	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

Tabla 5. Valores críticos de la prueba de F máxima de Hartley para homogeneidad de varianzas

Esta tabla brinda los valores críticos de la prueba del cociente de F máximo de Hartley para homogeneidad de varianzas cuando se tienen muestras de igual tamaño en cada grupo. Los valores críticos son dados para $\alpha = 0.05$, 0.01 (en negrito); el número de grupos p varía de 2 a 12; y v representa el número de grados de libertad para la varianza estimada.

		Número de grupos (p)										
v	α	2	3	4	5	6	7	8	9	10	11	12
2	0.05	39.0	87.5	142	202	266	333	403	475	550	626	704
	0.01	**199**	**448**	**729**	**1036**	**1362**	**1705**	**2063**	**2432**	**2813**	**3204**	**3605**
3	0.05	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
	0.01	**47.5**	**85**	**120**	**151**	**184**	**216**	**249**	**281**	**310**	**337**	**361**
4	0.05	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
	0.01	**23.2**	**37**	**49**	**59**	**69**	**79**	**89**	**97**	**106**	**113**	**120**
5	0.05	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
	0.01	**14.9**	**22**	**28**	**33**	**38**	**42**	**46**	**50**	**54**	**57**	**60**
6	0.05	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
	0.01	**11.1**	**15.5**	**19.1**	**22**	**25**	**27**	**30**	**32**	**34**	**36**	**37**
7	0.05	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
	0.01	**8.89**	**12.1**	**14.5**	**16.5**	**18.4**	**20**	**22**	**23**	**24**	**26**	**27**
8	0.05	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
	0.01	**7.50**	**9.90**	**11.7**	**13.2**	**14.5**	**15.8**	**16.9**	**17.9**	**18.9**	**19.8**	**21**
9	0.05	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
	0.01	**6.54**	**8.5**	**9.9**	**11.1**	**12.1**	**13.1**	**13.9**	**14.7**	**15.3**	**16.0**	**16.6**
10	0.05	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
	0.01	**5.85**	**7.4**	**8.6**	**9.6**	**10.4**	**11.1**	**11.8**	**12.4**	**12.9**	**13.4**	**13.9**
12	0.05	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
	0.01	**4.91**	**6.1**	**6.9**	**7.6**	**8.2**	**8.7**	**9.1**	**9.5**	**9.9**	**10.2**	**10.6**
15	0.05	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
	0.01	**4.07**	**4.9**	**5.5**	**6.0**	**6.4**	**6.7**	**7.1**	**7.3**	**7.5**	**7.8**	**8.0**
20	0.05	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
	0.01	**3.32**	**3.8**	**4.3**	**4.6**	**4.9**	**5.1**	**5.3**	**5.5**	**5.6**	**5.8**	**5.9**
30	0.05	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
	0.01	**2.63**	**3.0**	**3.3**	**3.4**	**3.6**	**3.7**	**3.8**	**3.9**	**4.0**	**4.1**	**4.2**
60	0.05	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
	0.01	**1.96**	**2.2**	**2.3**	**2.4**	**2.4**	**2.5**	**2.5**	**2.6**	**2.6**	**2.7**	**2.7**
∞	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.01	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Tabla 6 Área correspondiente al extremo derecho de una distribución Jí-cuadrada.

Gl	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.00004	0.00016	0.00098	0.00393	0.01579	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	15.98718	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	3.81575	4.57481	5.57778	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	6.30380	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	7.04150	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	7.78953	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	8.54676	22.30713	24.99579	27.48839	30.57791	32.80132
16	5.14221	5.81221	6.90766	7.96165	9.31224	23.54183	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	7.56419	8.67176	10.08519	24.76904	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	8.23075	9.39046	10.86494	25.98942	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	8.90652	10.11701	11.65091	27.20357	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	9.59078	10.85081	12.44261	28.41198	31.41043	34.16961	37.56623	39.99685
21	8.03365	8.89720	10.28290	11.59131	13.23960	29.61509	32.67057	35.47888	38.93217	41.40106
22	8.64272	9.54249	10.98232	12.33801	14.04149	30.81328	33.92444	36.78071	40.28936	42.79565
23	9.26042	10.19572	11.68855	13.09051	14.84796	32.00690	35.17246	38.07563	41.63840	44.18128
24	9.88623	10.85636	12.40115	13.84843	15.65868	33.19624	36.41503	39.36408	42.97982	45.55851
25	10.51965	11.52398	13.11972	14.61141	16.47341	34.38159	37.65248	40.64647	44.31410	46.92789
26	11.16024	12.19815	13.84391	15.37916	17.29189	35.56317	38.88514	41.92317	45.64168	48.28988
27	11.80759	12.87850	14.57338	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294	49.64492
28	12.46134	13.56471	15.30786	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824	50.99338
29	13.12115	14.25645	16.04707	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788	52.33562
30	13.78672	14.95346	16.79077	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218	53.67196
40	20.70654	22.16426	24.43304	26.50930	29.05052	51.80506	55.75848	59.34171	63.69074	66.76596
50	27.99075	29.70668	32.35736	34.76425	37.68865	63.16712	67.50481	71.42020	76.15389	79.48998
60	35.53449	37.48485	40.48175	43.18796	46.45889	74.39701	79.08194	83.29768	88.37942	91.95170
70	43.27518	45.44172	48.75757	51.73928	55.32894	85.52704	90.53123	95.02318	100.42518	104.21490
80	51.17193	53.54008	57.15317	60.39148	64.27785	96.57820	101.87947	106.62857	112.32879	116.32106
90	59.19630	61.75408	65.64662	69.12603	73.29109	107.56501	113.14527	118.13589	124.11632	128.29894
100	67.32756	70.06490	74.22193	77.92947	82.35814	118.49800	124.34211	129.56120	135.80672	140.16949